ADDIS ABABA UNIVERSITY

ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF CIVIL AND ENVIRONMENTAL ENGINEERING

Regionalization of Mean Annual Flow for Ungauged Catchments (Case study: Abbay River Basin)

A Thesis in Hydraulic Engineering

By Habtamu Nega

A Thesis

Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science

October. 2019

Addis Ababa

The undersigned have examined the thesis entitled '**Regionalization of Mean Annual Flow for Ungauged Catchments (Case study: Abbay River Basin)'** presented by **Habtamu Nega**, a candidate for the degree of **Master of Science** and hereby certify that it is worthy of acceptance.

______________________	________________	__________
Advisor	Signature	Date
______________________	________________	__________
Internal Examiner	Signature	Date
______________________	________________	__________
External Examiner	Signature	Date
______________________	________________	__________
Chairperson	Signature	Date

UNDERTAKING

I certify that research work titled "**Regionalization of Mean Annual Flow for Ungauged Catchments**" is my own work. The work has not been presented elsewhere for assessment. Where material has been used from other sources it has been properly acknowledged/referred.

Habtamu Nega

Email: Habtamushafi@gmail.com

ACKNOWLEDGMENTS

I would first like to thank my thesis advisor Yilma Seleshi (Ph.D.) for his genuine support starting from the initiation of the thesis idea up to the end. His office was always open whenever I ran into a trouble spot or had a question about my research or writing. He consistently allowed this paper to be my own work but steered me in the right direction whenever he thought I needed it.

I would also like to acknowledge the staff members MoWIE in the Hydrology department, librarians and GIS department and NMA specially Zeki in their positive and inspiring hospitality whenever I run any support from them. Additionally, my appreciation goes to the staff members of the School of Civil and Environmental Engineering at Addis Ababa University whenever I want any support throughout my stay.

Finally, I must express my very profound gratitude to my parents and to my left ribs, classmates, and friends for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them.

ABSTRACT

The future existence and development of potential water resources hinge mainly on the past features of hydrological behaviors. Unfortunately, these features are not available depending on the socio-economic and political stability of the country. Streamflow measurement is one of the features of our country Ethiopia and other developed and developing countries suffer to estimate. This study aims to identify the rainfall-runoff relationships and develop an empirical equation that can be used to estimate the mean annual flow for ungauged catchments in the Abbay river basin.

Based on the influence on runoff generation and easy availability climate variable (rainfall) and physiographic variables (catchment area, land slope, and elevation) are selected as a predictor of flow in the Abbay river basin. The selected 27 hydrometric stations split into training (70%) and test (30%) datasets for the periods 1981 to 2013. Linear and non-linear forms of the equation were tested to relate annual flow and annual rainfall at each hydrometric station. A regression model used to relate the dependent variable mean annual flow and independent variables climate and physiographic characteristics. Statistical tests used for selecting the best trustworthy equation.

Analysis of the relationship between runoff and rainfall demonstrates fully dominance of non-linear forms of equations and development of an empirical equation shows that the catchment area, mean annual rainfall, and the average elevation is the trustworthy model with both R^2 and Ns value of 0.96. Furthermore, the validation analysis strengthens this equation statistically and graphically. The result reveals that relationships of rainfall and runoff are non-linear and the size of the catchment, rainfall, and elevation of the catchment highly governs the amount of flow expected at the outlet. On this root, it is recommended that interested professionals on the estimation of flow at ungauged catchments carefully delineate the area, estimate rainfall with enough rain gauge stations, and consider elevation as the factor. Moreover, further studies need to incorporate additional catchment characteristics to identify other influential factors in the formation of runoff.

Key Words: Rainfall-Runoff relationships, Ungauged Catchments, Regionalization, Regression Analysis, Mean Annual Flow

ABBREVIATIONS

AE	Average Elevation
AS	Average Slope
DEM	Digital Elevation Model
GIS	Geographic Information System
HBV	Hydrologiska Byråns Vattenbalansavdelning
IAHS	International Association of Hydrological Sciences
MAF	Mean Annual Flow
MAR	Mean Annual Rainfall
MoWIE	Ministry of Water, Irrigation, and Energy
NWSRFS	National Weather Service River Forecast System
OLS	Ordinary Least Squares
SHE	Système Hydrologique Européen
VIF	Variance Inflation Factor

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 Background

One of the world's most important and vital natural resources is water (Anisa Z. and Amanda E., 2018). Water is one of human being basic needs that has its own involvement in day-to-day activities all over the world. Water resource planning and projects have been implemented for centuries to accomplish the need of the people in the way of electric power generation, domestic water use, irrigation, recreation as well as transportation. However, these purposes of water resources basically require hydrological data.

Hydrological data are collected from gauge stations either hydrometric or mereological stations. But, most of the catchments are ungauged and even gauged catchments are not having sufficient recorded length of required data.

The above-mentioned problem of ungauged catchment or poorly ungauged catchment is including both developed and developing countries all over the world. However, our country Ethiopia is one the developing country which has huge water resources in Africa having 12 basins that affected by a number of ungauged catchments that can be used for a different purpose and add economic development in the country. As indicated by (Chekole T. and Abdella K., 2016) the situation in Ethiopia is problematic, as there are no evenly distributed hydrometric stations, as large areas lack gauging stations; and only a few years of data are available.

One of the hydrological data that always used for planning and design of water resource projects is streamflow data. This shows that there must be another way to use water resources in ungauged catchments to gain streamflow data. A regional regression model is one of the most common adapted methods of estimating runoff (Anisa Z. and Amanda E., 2018; Vogel et al, 1999; Barbarossa et al, 2017; Burgers et al, 2013; D. M. Papamichail et al, 2002).

The aim of the research is to identify the rainfall-runoff relationship and develop a regional empirical equation for ungauged catchments using gauged catchments in the Abbay river basin of Ethiopia using regression models that consider different characteristics of catchments.

1.2 Statement of the problem

Water is needed for different activities that are vital for the existence and development of human society (S.K. Jain and V.P. Sing, 2003). Water resources for water supply, hydropower plants, irrigation, transport, and recreation are the main purpose that human practice all over the world. Moreover, the water resource in the world is its own effort on life and the existence of wildlife and ecosystem. Eventually, water has a huge devastating effect that will damage property and the loss of the irreplaceable life of humans by flood, drought, and waves in the oceans. This usefulness and consequence of water in the world to some degree controlled by the design and management of different water resource schemes.

The design and management of hydraulic structures is done by the help of hydro-meteorological gauges which are record different hydrological information in the basins such as streamflow, precipitation and water level, etc. However, coverage of hydro-meteorological gauging stations is a vast problem all over the world, especially in developing countries like Ethiopia; which is the tower of water resources in east Africa composed of 12 river basins. Additionally, insufficient record period is also a problem even in gauged catchments. According to (Chekole T. and Abdella K., 2016) the situation in Ethiopia is problematic, as there are no evenly distributed hydrometric stations, as large areas lack gauging stations; and only a few years of data are available.

Tara R. and Paulin C., (2013) and M. Sivapalan et al., (2003) stated that ungauged catchment or poorly gauged catchment definition depends on the variable from the catchment we want to practice. M. Goswami et al., (2007) mentioned that “Lack of data, both qualitative as well as quantitative, often inhibits the undertaking of scientific analyses for such catchments which are required for purposes such as assessing the water resource, ensuring its long term availability, forecasting its occurrence over short lead-times, predicting its future occurrence, and developing its source”. The essential flow of the river is one of the requirements in the assessment of water resource schemes in time and space.

Yusuf Javeed and Apoorva KV (2015) indicated that streamflow measurements are important to obtain climate and/or land-use change for future assessment of the hydrological behavior of river basins. In order to obtain responsible management, estimates of annual watershed runoff volumes are required (Vogel et al., 1999).

MoWIE Research and Development Directorate, (2018) identified as a researchable issue in surface and groundwater hydrology, that most of the catchments in the country are ungauged: as a result, they suggested the development of improved tool and method for estimation of flow for ungauged catchment is needed.

Planning and design of water resource projects in an ungauged catchment are solved by regional analysis (Yu et.al, 2009). Regionalization states the transferring of hydrologic information from gauged catchment to ungauged catchment for hydrological homogeneous regions.

The regression model equation relates streamflow and catchment characteristics in an explanatory way that helps to understand general hydrological patterns and processes adopted across different scales. Global-scale (Barbarossa et al, 2017; Burgers et al, 2013) and regional scale (Anisa Z. and Amanda E., 2018; Vogel et al., 1999; Sedighi, 2008; D. M. Papamichail et al., 2002; Tran et al., 2015) regression model approach for mean annual flow using catchments characteristics are developed. However, in Ethiopia, the regionalization study is focused on flood frequency Analysis (Abebe S.G. et al, 2013; Gebeyehu, 1989) and using a rainfall-runoff model (Chekole T. and Abdella K. 2016; Mesgana B., 2013; Tesfaye, 2011).

Therefore, an attempt is done to identify relationships between rainfall and runoff and formulate an empirical equation for the Abbay river basin using regression models.

1.3 Objectives

1.3.1 General objective

The main objective of this research is to establish a way that used to estimate mean annual flow for an ungauged catchment in the Abbay river basin.

1.3.2 Specific objective

- To assess relationships between rainfall and runoff in the basin
- To generate an empirical regression equation for ungauged catchments

1.4 Research Questions

- What is the relationship between rainfall and runoff in the Abbay river basin?
- Which catchment characteristics should be included in the model and what criteria are used to identify the model trustworthiness?

1.5 Significance of the study

This study is very important for the country to develop most of the potential water resource projects in the basin. The problems that can't be answered due to the unrecorded river that can be used for different purposes in the sufficient and effective ways for the society in the nearby area as well as to the national level possibly removed by this thesis.

Moreover, for those professionals who suffer due to ways of estimating flows in ungauged and poorly gauged catchments for the development of water-related projects this thesis may help. Furthermore, this thesis helps to understand the behavior or characteristics of the catchments that makes confused.

1.6 Scope of the study

The purpose of this study is to identify the relationship between rainfall-runoff and to develop an empirical equation that can be used to estimate mean annual flow for ungauged catchments in the Abbay river basin. Regression models that relate mean annual flow and four catchment characteristics excluding geomorphologic and anthropogenic variables used. The climatic variables (mean annual flow and mean annual rainfall) for the period 1981 to 2013 used. For the selection of best model equation coefficient of determination R^2 including adjusted, Nash Sutcliffe and verification of the equation by nine hydrometric stations that were not included in the development of the equation.

CHAPTER 2 LITERATURE REVIEW

2.1 General

Water is essential to life. The purpose of water is in diverse ways to human life, first of all, it's one of the crucial needs, residential use, agricultural use, power generation, navigation and recreation, and also used as a source of foods like fish. Additionally, water is useful for flora and fauna. Water, in excess or in absence, is the greatest problem the world faces for a century. Human activities disrupt the natural hydrological and ecological regimes of river basins across the globe which leads to a shortage of water supply for the people, consecutive and highly distractive flood and loss of biodiversity (M. Sivapalan et al., 2003).

Water resource development projects are designed and implemented for purposes revealed earlier. However, the change of rainwater into a runoff process is quite complex. Hence, the relationship between rainfall and runoff will add a solution at least for that specific area studied. Additionally, the amount of flow passing through the rivers is not known, even though there is plenty of a number of gauged stations on the river. Regional analysis is one of the methods of determining the unknown amount of flow using the recorded amount of flow of rivers in that basin. Space and time variability of mean annual flow information is required for the use and development of water resources (Waylen.P. and M. Woo, 1982).

2.2 Rainfall-runoff relationships

A rainfall-runoff relationship can be carried out within a purely analytical framework based on observations of the inputs and outputs from a catchment area (Bhagat NK, 2017). However, the rainfall-runoff process is complex, dynamic, and nonlinear phenomena governed by different known and unknown physiographic attributes of the catchments that show Spatio-temporal variability. The rainwater experiences a number of transformation and abstraction through falling to the earth and finally goes to the catchment outlet as runoff. The collection of these data is difficult, time-consuming, and also expensive.

Determining a strong relationship between rainfall and runoff is one of the problems in water resources. This problem solved using rainfall-runoff models. Based on model input and parameters and the extent of physical principles applied in the model the general classification of the rainfall-runoff model into three categories empirical, conceptual, and physical model. Furthermore, based on the spatial and temporal understanding of the model's catchment area can be classified as lumped, semi-distributed, and distributed models (Devi et al., 2015; Jan Sitterson et al., 2017). According to Vaze J. et al., (2012) the rainfall-runoff model approaches in order of simple to complex are empirical methods, large scale energy-water balance equations, conceptual rainfall-runoff models, landscape daily hydrological models, and fully distributed physically-based hydrological models which explicitly model hillslope and catchment processes.

The empirical model is a data-driven model that uses nonlinear statistical relationships between inputs and output time series and not from the physical process of the catchment. These are observation oriented models that take only the information from the existing data without considering the features and processes of the hydrological system. Mostly adapted empirical models used are regression equation, Machine learning ANN, and SCS-curve number. This model uses one up to five number of parameters with a low risk of overfitting and best use for ungauged watershed if a runoff is the only output.

The conceptual model defines all of the simplified components of hydrological processes and connects each other using a series of mathematical equations. It consists of a number of interconnected reservoirs (Buckets) or conceptual stores that represent the physical elements in a catchment (Devi et al., 2015; Vaze J. et al., 2012). Lack of physical meaning in governing equations and parameters is also a limitation. Conceptual models are best used when computation time is limited and catchment characteristics are not analyzed in detail. Examples of conceptual models are TOPMODEL, HBV, NWSRFS, SIMHYDand HSPF. This model easy to calibrate and uses two up to four numbers of parameters with a very low risk of overfitting.

The physical model also called mechanistic models which are based on the understanding of the physics related to the hydrological processes. It uses state variables that are measurable and are functions of both time and space. A discretization of spatial and temporal coordinates is made at a very fine scale for the entire catchment and the physical equations are solved for each discretized grid to obtain a solution. The general physics laws and principles used include water balance equations, conservation of mass and energy, momentum, and kinematics. St. Venant, Boussinesq's, Darcy, and Richard's are some of the equations adopted by physical models (Pechlivanidis et al., 2011 cited in Jan Sitterson et al., 2017). Some examples of physical models are VIC, SHE/MIKE SHE. This model is site-specific, needs large numbers of parameters ten to thousand's, and best use for small scale with great data availability.

The choice of an appropriate model depends on the purpose of the modeling, nature of the system to be modeled, the hydrological element(s) to be modeled, availability of input data, the applicability of the model, and the accuracy of the output (Vaze J et al.., 2012).

2.3 Ungauged catchments

Streamflow is one of the crucial data for hydrological applications in such a way as water resource system planning and management, disaster risk management, and environmental impact assessment. Understanding of runoff generation mechanism and estimation of the amount of runoff for water resources system planning and management depends on spatial location and temporal resolution of observed hydrological and climatic data (Zelelew, 2012). Depending on the interest of information needed in the catchment it's possible to say ungauged (Tara R. and Paulin C., 2013; M. Sivapalan et al., 2003); though, there is also station which is less length of record usually called poorly gauged.

Estimation of runoff in ungauged catchments was one of the problem hydrologists faces for years. The estimation starts with developing unit hydrograph which has been widely and successfully used over the past decades. Unit hydrograph firstly announced by Sherman in 1932 resulting from one-unit (1 inch or 1 cm) rainfall excess uniformly distributed spatially and temporally over the catchment for the entire specified duration (Chow et al., 1988). However, the unit hydrograph developed from rainfall and streamflow data only used for a watershed that data measured. So, synthetic unit hydrograph was developed to use unit hydrograph for the other location on the stream in the same watershed or for the nearby watershed that has the same characteristics. Thus, synthetic unit hydrograph is a method of developing unit hydrograph for ungauged catchments.

According to Chow et al., (1988) there are three types of synthetic unit hydrograph. These are:

i. Those relating hydrograph characteristics (peak flow rate, base time, etc.) to watershed characteristics (Snyder unit hydrograph, 1938),
ii. Those based on a dimensionless unit hydrograph (SCS unit hydrograph, 1972), and
iii. Those based on models of watershed storage (Clark unit hydrograph, 1945)

The first synthetic unit hydrograph introduced by Snyder on a study of watersheds located in the Appalachian Highlands and varying in size from 30 to 30,000Km^2. This synthetic unit hydrograph is based on relationships found between three characteristics of a standard unit hydrograph (effective rainfall duration, t_r, the peak direct runoff rate, q_p, and the basin lag time, t_p) and descriptors of basin morphology. From these relationships, five characteristics of a required unit hydrograph for a given effective rainfall duration may be calculated the peak discharge per unit of watershed area, q_{pR}, the basin lag, t_{pR}, the base time, t_b, and the widths, W (in time units) of the unit hydrograph at 50 and 75 percent of the peak discharge (Chow et al., 1988).

Snyder defines standard unit hydrograph by relating effective rainfall duration,t_r, and basin lag t_p by $t_p = 5.5*t_r$ and for the standard unit hydrograph the basin lag, t_p, and the peak discharge, q_p, are given by $t_p = C_1 C_t \, (L * L_c)^{0.3}$ and $q_p = \frac{C_1 C_p}{t_p}$ respectively. Where: The basin lag time is in hours, L is the length of the mainstream in kilometers (miles), Lc is the distance in kilometers (miles), and $C_1 = 0.75$ (1.0 for English units). C_t and C_p is a coefficient derived from gauged watersheds in the same region. The peak discharge of the standard unit hydrograph is in m^3/s (cfs), and $C_2 = 2.75$ (640 for English units). From a derived unit hydrograph of the watershed, values of its associated effective duration t_R in hours, its basin lag t_{pR} in hours, and its peak discharge q_{pR} in m3/s are obtained. If $t_{pR} = 5.5t_R$, then the derived unit hydrograph is a standard unit hydrograph and $t_r = t_R$, $t_p = t_{pR}$, and $q_p = q_{pR}$, and C_t and C_p are computed by the equations for t_p and q_p given above, corresponding to the standard unit hydrograph. If t_{pR} is quite different from $5.5t_R$, the standard basin lag is computed using:

$$t_p = t_{pR} + \frac{t_r - t_R}{4}$$

This equation must be solved simultaneously with the equation for the standard unit hydrograph lag time, $t_p = 5.5tr$, in order to obtain t_r and t_p. The value of C_t is then obtained using the equation for t_p corresponding to the standard unit hydrograph. The value of C_p is obtained using the expression for q_p corresponding to the standard unit hydrograph, but using $q_p = q_{pR}$ and $t_p = t_{pR}$. When an ungauged watershed appears to be similar to a gauged watershed, the coefficients C_t and C_p for the gauged watershed can be used in the above equations to derive the required synthetic unit hydrograph for the ungauged watershed

After that required synthetic unit hydrograph related to standard unit hydrograph by the following equation

$$q_{pR} = \frac{q_p t_p}{t_{pR}}$$

Assuming a triangular shape for the unit hydrograph, and given that the unit hydrograph represents a direct runoff volume of 1 cm (1 in), the base time of the required unit hydrograph may be estimated by, $t_b = \frac{C_3}{q_{pR}}$ where $C_3 =$ 5.56 (1290 for the English system). Lastly, the width in hours of the unit hydrograph at a discharge equal to a certain percent of the peak discharge q_{pR} is given by $W_{\%} = C_w(q_{pR})^{-1.08}$. here: The constant C_w is 1.22 (440 for English units) for the 75% width and equal to 2.14 (770 for English units) for the 50% width. Usually, one-third of this width is distributed before the peak time and two-thirds after peak time. In this method of developing the unit hydrograph for ungauged catchments it is necessary to identify a nearby and physiographically similar gauged watershed from which to transfer parameters. Even if a gauged watershed is located near to the ungauged watershed of interest, the greatest limitation of the method is the difficulty in verifying that the two watersheds are sufficiently alike to transfer C_t and C_p from one to the other. However, in the absence of gauge data on both watersheds, it can never be known for sure that the transfer is proper.

The second method of developing synthetic unit hydrograph is SCS dimensionless hydrograph developed by Victor Mockus and was derived based on a large number of unit hydrographs from basins that varied in characteristics such as size and geographic location (NOHRSC, 2019). In this method discharge expressed by the ratio of discharge q to peak discharge q_p (ordinate) and the time by the ratio of time *t* to the time of rising of the unit hydrograph, T_p (abscissa). Given the peak discharge and lag time for the duration of excess rainfall, the unit hydrograph can be estimated from the synthetic dimensionless hydrograph for the given watershed. The SCS suggests that the dimensionless unit hydrograph (DUH) can be described in terms of an equivalent Triangular Unit Hydrograph (TUH). The values of q_p and T_p can then be estimated using this simplified triangular unit hydrograph whose height is equal to q_p and whose time base,t_b, is equal to $2.67T_p$. The time is usually expressed in hours and the discharge in $m^3/s/cm$. After analysis of a great number of unit hydrographs, the SCS recommends a recession duration of $1.67T_p$. Because the volume of direct runoff must equal 1 cm, it can be shown that $q_p = \frac{CA}{t_p}$, where C = 2.08 (483.4 in the English system) and A is the drainage area in square kilometers (square miles). From a study of many large and small rural watersheds the basin lag is $t_p = 0.6Tc$, where Tc is the time of concentration of the watershed. The time to peak, Tp, is then expressed as $T_p = \frac{t_r}{2} + t_p$, where t_r is effective rainfall duration. However, this SCS dimensionless unit hydrograph or triangular unit hydrograph provides an only empirical approximation of flood runoff characteristics; its reliability is limited to the type and the size of the catchments which were used for its derivation.

The last method of the synthetic unit hydrograph is developed by Clark in 1945 that combined a parameter to model the watershed storage and time of concentration. Clark noted the translation of flow through the watershed was described by a time-area curve that expresses the fraction of watershed area contributing runoff to the watershed outlet as a function of time since the start of effective precipitation. A linear reservoir was used by Clark to reflect the storage affects of watersheds. Clark only used gauged catchments in his original work and did not provide guidelines for the estimation or determination of the R-value and time-area relationships for ungauged catchments. It is possible to transfer the R-value from one catchment to another nearby catchment through regression analysis. Parameters to be considered (but not limited to) are drainage area, lengths, and slopes. The U.S.Army Corps of Engineers, Hydrologic Engineering Center, has noted that the ratio storage coefficient (R), sum of time concentration (t_c) and storage coefficient $R/(t_c+R)$ tends to remain constant for a region. In this method Clark did not consider ungauged catchments in his original work, his method of estimating R relies upon gauge information. The transfer of the model parameter to ungauged catchments using regression analysis at a regional scale requires a number of gauged catchments to develop the equations to determine the Clark parameters, but this would be difficult in regions where most of the catchments are ungauged.

P.K. Singh et al., (2014) review synthetic unit hydrograph by grouping into four, 1) traditional or empirical model; 2) conceptual models; 3) probabilistic models, and 4) geomorphological models. The traditional models are based on different empirical equations and have certain region-specific constants/coefficients varying over a wide range. In this method of development of synthetic unit hydrograph (SUH) some degree of subjectivity is involved in fitting the remaining points. In addition, simultaneous adjustments are also required to ensure that the area under the SUH is unity corresponding to unit rainfall-excess. These models are the most widely used in engineering problems even though it has a limitation of inconsistences. In this review journal three methods Snyder (1938), Taylor & Schwarz, and Soil conservation service (1957) discussed as an example in this method.

The conceptual models are based on the continuity equations and linear storages discharge relationship. In this group Clark IUH model (Clark, 1945), Nash IUH model (Nash, 1957), Dooge IUH model (Dooge, 1959), Nonlinear IUH model (Singh, 1964), Urban parallel cascade IUH model (Disken et al., 1978), Hybrid model (Bhunya et al., 2005), and Extended hybrid model (Singh et al., 2007) are the models that are discussed in this review. The third group of the review probabilistic or probability distribution function based model uses a parametric approach and employ the density function for SUH derivation. Grey's method (Gray, 1961), Croley method (Croley, 1980), Transmutation approach (Singh, 2000), and Simplified gamma approach (Bhunya et al., 2003) are the model's reviewed in this group. Lastly, the geomorphological group uses basin geomorphology to develop instantaneous unit hydrograph. Here, the geomorphological instantaneous unit hydrograph (GIHU) model, GIUH-based 2PGD Nash model, and width function-based GIUH model are examples of models discussed in this group. Additionally, the application of the digital elevation model (DEM) in the geomorphological instantaneous unit hydrograph was also discussed. Finally, a summarization of important studies in the synthetic unit hydrograph that was incorporated in the review.

The International Association of Hydrological Sciences (IAHS) has launched a long-term research initiative by declaring the years 2003 – 2012 as the IAHS Decade on Predictions in Ungauged Basins (PUB), which is commonly referred to in hydrology community as the PUB initiative. The aim of this initiative was to "formulating and implementing appropriate science programs to engage and energize the scientific community, in a coordinated manner, towards achieving major advances in the capacity to make reliable predictions in ungauged basins." (M. Sivapalan et al., 2003).

According to M. Sivapalan et al., (2003) methods appropriate for ungauged basins include extrapolation of response information from gauged to ungauged basins, measurements by remote sensing, application of process-based hydrological models where the climate inputs are specified or are measured, and application of combined meteorological-hydrological models without the need to specify precipitation inputs.

In Ethiopia, the spatial and temporal distribution of hydro-metrological station is not enough (Chekole T. and Abdella K., 2016). Even though, the available gauged stations are suffering large gaps in the record. For ungauged catchments of Didessa sub-basin, Chekole T. and Abdella K., (2016) a semi-distributed conceptual hydrological model HBV-96 was used for prediction of discharge. Seven gauging stations were used for the study but only four of them were capable of performed better. Mesgana B., (2013) estimates monthly runoff of ungauged catchments in the Baro-Akobo basin for his MSc thesis using the WATBAL conceptual lumped rainfall-runoff model.

2.4 Regional Analysis

Regionalization methods intend to estimate an unknown variable by either inferring it from physiographic catchment descriptors or combining the values from similar catchments or neighboring ones (J. Odry and P. Arnaud, 2017). In other words, regionalization analysis conducted mostly in such a way that there is a similarity in catchments in view of the distance between them and their similarities in the catchment and physiographic characteristics. For the first approach of regionalization analysis which is spatial proximity, the main idea is that catchments in neighbors will have the same climate and catchment features. Therefore, a particular model approach and accompanying calibrated model parameter values from gauged catchments can be derived and applied at the ungauged catchments in order to predict the discharge. The later conventional approach of regionalization consists of three steps. The first step implies calibration of the chosen model structure for a large number of catchments for which sufficiently long and informative observations of discharge are available. This can be done using several criteria which are to be established. Secondly, an attempt is made to derive regression equations, which is the most commonly used method, which predicts the model parameter values using one or a combination of catchment characteristics. Finally, parameter values for the ungauged catchment can be estimated using the regional model, which comprehends all the established relationships between catchment characteristics and model parameters merged in the hydrological model, and a runoff prediction can be made.

Regional analysis was adopted in hydrology for numerous studies such as flood frequency analysis (Gebeyehu, 1989; J. Odry and P. Arnaud, 2017; Abebe S.G. et al., 2013), streamflow prediction (Anisa Z. and Amanda E., 2018; M. Goswami et al., 2007; Pasquale C. et al., 2006; Vogel et al., 1999).

Tara R. and Paulin C., (2013) review regionalization methods that are used for streamflow prediction in ungauged basins from period1990 to 2011. They define regionalization as a process of transferring hydrological information from gauged to ungauged or poorly gauged basins to estimate streamflow. Accordingly, regionalization requires information of catchment attributes (physiographic and meteorological attribute) and a function (linear or nonlinear) for relating predictors to the predictand. In addition they stated five main important steps in regionalization analysis; the first and second steps are collect and manage catchment attributes, which are metrological attributes (such as mean annual rainfall and temperature) and physiographic information (such as catchment area, soil type, area covered by grass, trees) and determination and classification of hydrological variables of interest respectively. Development of a relationship between streamflow or hydrological model parameters and catchment attributes and evaluation of model performance by using pseudo-ungauged basins before applying the model are the third and fourth steps. Finally, incorporating uncertainty analysis.

Furthermore, Tara R. and Paulin C. (2013) discuss catchment attributes used in different studies and they realize that an initial hypothetical judgment is required to identify which potential catchment attributes would have an impact on the runoff responses of interests. Their assessment of previous studies is that physiographic information (catchment area, elevation, and basin slope) and meteorological attribute (mean annual or daily rainfall and temperature) are most often combined in regional analysis. The percentage of area covered by water or by land use, location of stream gauges or centroid of catchments, mean annual evaporation and snowfalls are other infrequent combined catchment attributes in different studies. Generally, they conclude that the catchment area, the slope of the basin/channels/, and mean annual or daily rainfall and temperatures are the most widely used attributes by researchers.

Regional analysis is stated as one of the approaches which are appropriate for the prediction of ungauged basins (PUB) in the International Association of Hydrological Sciences (IAHS) (M. Sivapalan et al., 2003).

In Ethiopia Gebeyehu, (1989) develop the regional equation of flood frequency analysis for Ethiopia taking 78 catchments using the Index-flood method of regional analysis for flood frequency analysis. Chekole T. and Abdella K., (2016) in their study of estimation of discharge for ungauged basin in Didessa sub-basin, four regionalization methods (Spatial proximity, Area ratio method, Sub-basin mean method and regional model) were tested with observed discharges as a result regional method of regionalization were selected as best for prediction of discharge and spatial proximity were as least method. Likewise, Tesfaye, (2011) in his MSc thesis titled predicting discharge at ungauged catchments using rainfall-runoff model on Omo-Gibe river basin also adopted the same four regionalization methods and select regression analysis as the best method among other and area ratio method generally performs slightly better than the sub-basin mean and the spatial proximity method performs the least. Additionally, Berhane, (2013) also uses the WatBal rainfall-runoff model to estimate monthly flow in ungauged catchments on the Baro-Akobo river basin and regression analysis was used to transfer model parameters.

Gebeyehu H., (2013) select sub-basin means and regression methods of regionalization to estimate model parameters of HBV-96 semi-distributed conceptual models for ungauged catchments in the Awash river basin. Validation of the model shows a regression analysis method performed moderately well (Ns = 0.64) and the sub-basin mean method performs unsatisfactory (Ns – 0.25). Abebe S.G. et al., (2013) conduct regional analysis for the Abbay basin firstly by identifying five homogeneous regions using L-moment skewness (L_{Cs}) and L-moment kurtosis (L_{Ck}). After that index flood method of regionalization for ungauged catchment was done.

Regionalization approaches generally can be divided into two as hydrologic model-dependent and hydrologic model-independent. The first one is about the transferring of rainfall-runoff model parameters between basins. However, for the second approach, the equation and its parameters are transferred (Tara R. and Paulin C., 2013). Among the hydrologic model-independent methods for streamflow regionalization (regression-based, time-series model, and scaling relationships) regression equations (including linear and nonlinear) developed between the desired hydrologic responses and catchment attributes are the most commonly used ones. One of the advantages of hydrologic model-independent methods is the lower data requirement of these approaches and the simplicity of their structure that does not require special knowledge and expertise of hydrological modeling. However, the identification of an appropriate model structure (e.g., ANN architecture) requires some expertise. Data-driven methods do not simulate the actual rainfall-runoff process, and therefore, they are not impacted by uncertainty because of the physical process being modeled; however, they are still affected by other sources of uncertainty, e.g., the estimation method and its parameterization.

Ralf M. et al., (2006) review spatial proximity and catchment attributes regionalization methods for large catchment size and concluded that there is no single best regionalization method and the results of some of the studies are seemingly conflicting.

In the Abbay river basin master plan study (BCEOM, 1999) regional analysis of flood discharge for 100, 1000, and 10000 years of return periods was done to fit regression analysis using catchment area as the only predictor by dividing it as less than 10000Km^2 and greater than 10000Km^2. Furthermore, in the study, an attempt was made to relate mean slope with flood discharge but the relationship was not acceptable or shows less R^2.

Genale Dawa river basin integrated resources development master plan study (LAHMEYER, 2007) regional analysis of mean annual flood was done for all the basin and sub-basin level using area and mean streamflow. In the basin level mean annual flood (m^3/s) and Area (Km^2) related with power rule and gave R^2 of 0.71 and relating mean annual flood (m^3/s) and mean streamflow (m^3/s) R^2 become 0.88. Furthermore, an attempt was made to estimate mean flow at ungauged catchments by nominating a key hydrometric station on the river of interest and calculating runoff for the incremental area, upstream or downstream of the site, according to mean elevation of the incremental area.

2.5 Regression model

According to Mauro Naghettini et al., (2017) regression analysis includes two-step procedures. First step preparation of function between predictand and predictor and estimation of numerical values of the regression coefficient. The second step is the evaluation of the regression model. Regression model in this case relate mean annual flow with catchment and physiographic characteristics which have been widely used in the world global scale (Barbarossa et.al, 2017; Burgers et.al, 2013) and region-based (Anisa Z. and Amanda E., 2018; Vogel et al., 1999; Sedighi, 2008; D. M. Papamichail, et al., 2002; Tran et al., 2015; Gebeyehu, 1989).

Anisa Z. and Amanda E., (2018) conduct their MSc thesis on regional regression models of MAF and design flow in a tropical region of Colombia. Influential variables are established using principal components analysis (PCA) and stepwise regression is done then by k-means clustering and Andrews curves they identify hydrologically homogeneous regions (HHR). They select geomorphologic variables (drainage area, the perimeter of basin, hydraulic radius of the basin and mean basin elevation) and climatic variables (mean temperature, mean evaporation and mainly mean precipitation for different months). Likewise, other mentioned studies' best suitable regression model is selected by starting with one predictor and added other predictors by testing with different statistical tests p-value, residual histogram, VIF, R^2 and residual scatter plot.

Barbarossa et al, (2017) done a global-scale regression model on 1885 catchments which makes them different from (Burgers et al, 2013) is that they added more climatic and physiographic variables such as mean annual temperature, mean annual slope and mean annual elevation variables for their model. Generally, they used 2 climatic variables (precipitation and temperature) and 3 physiographic variables (Area, slope, and elevation). Akaike information criterion (AIC) and Bayesian information criterion (BIC) are used as a statistical test to identify a parsimonious model for individually established predictors. Furthermore, to make free from biased coefficients of the regression model they used Cook D influence statistics. Finally, they test their model by first establishing one predictor variable, then add one predictor variable until the five predictors are set and they concluded that a full set of predictors shows the most parsimonious model (Low AIC) which shows 89% variation in MAF. They also put the order of predictors by giving catchment areas as the most important and followed by precipitation, temperature, slope, and elevation. Their model performs better for higher MAF values and residual error slightly larger for catchments with lower precipitation values and at higher altitudes.

Tran et al, (2015) used the regression model for the upper Mississippi river basin for climatic, geomorphologic and anthropogenic variables as predictors of mean annual streamflow. What makes them different from the other is that they developed the model by a spatial statistical approach. Nash-Sutcliffe model efficiency of 0.985 and R^2 of 0.993 was obtained from the model.

Additionally, Burgers et al, (2013), also conducted a regression model by taking 663 stations worldwide and assign each station in one of the five climate zones equatorial (A), arid (B), warm temperate (C), snow (D) and polar (E) at a global scale using average and maximum discharge, catchment area, river main-stem length, and precipitation. Ordinary least square and standardized major axis were used to drive the empirical equation. By taking the catchment area as the only predictor, the variance of a model in MAF gives 40% and also, they take precipitation as the only predictor, an 8% variation of a model in MAF. However, their model results of 56% of the variation in MAF by combining catchment area and precipitation which is a multiple regression model.

Sedighi, (2008) in his dissertation used regression analysis for estimation of Centroid lag and Centroid Lag-to-Peak by taking seven watershed characteristics (area, slope, stream density, basin roughness, compactness ratio, percentage of karst area, and percentage of impervious area) and rainfall characteristics (duration and volume) as predictors designed for 16 watersheds in Southwest Florida .

D. M. Papamichail et al, (2002) performed a multiple regression model for 11 catchments in Chalkidiki, Greece using the Water balance model. They took eight catchment characteristics, annual rainfall, mean annual temperature, catchment area, average land slope, mean elevation, length of the stream, average stream slop and length along the longest watercourse from point of concentration ratio to a point opposite the centroid of the drainage basin. They divided the recorded data into two groups as first group calibration data and second group validation data. The step-wise regression technique using SPSS statistical package was used for the development of the multiple regression model of eight predictors. Moreover, from the different developed alternatives, they selected the one that exhibited standard error of estimation (0.8044) and high R^2 (0.78) as well as satisfied least squares assumptions.

Vogel et al, (1999) developed a regression model on 1553 watersheds within different 18 regions in the USA. The first two moments (Mean and Variance) of mean annual flow are related as a function of regression model variables of climate and geomorphology. Variance inflation factor (VIF) was used to check multicollinearity and Cook D influence statistics used for biased coefficients of the regression model. Like the other studies, they started with drainage area as alone predictor and adjusted R^2 ranges 27.3% - 99.1% (mean 71.4%). Furthermore, mean annual precipitation and mean annual temperature are included in the drainage area as predictor adjusted R^2 improved to 85.2% - 99.7% (mean 94.5%).

Gebeyehu, (1989) used a regression model for the whole country's flood frequency analysis. In the first attempt, he took average annual rainfall, area, stream length, the slope of mainstream, shape index and elevation as predictor variables but the area and stream length show correlation (0.85) as well their log (0.92); as a result, stream length exclude from the variables. He used 62 catchments as an initial model firstly area as the only predictor which gives R^2 of 69%. However, the addition of average annual rainfall improves R^2 to 79%. Similarly, slope, shape, and elevation with the area as predictor variables don't exhibit improvement on the coefficient of determination R^2 of 70%, 69%, and 69% respectively. Moreover, adding of slope and elevation into area and average annual rainfall was gave him 1% and 3% upgrading of coefficient of determination R^2. So, he concluded that area and rainfall were taken for maximum discharge predictors by adding 16 catchments from the initial 62 catchments. Hence, for 78 catchments area alone as predictors 79% of R^2 with a standard error of 0.2585 obtained. Additionally, area with four forms of rainfall (average annual rainfall, annual maximum monthly rainfall, maximum measurements in two consecutive months, and maximum rainfall in three consecutive months) and he selected the one that exhibited minimum standard error (0.2171) and maximum R^2 (0.85).

Regression analysis was used to estimate mean annual flows at ungauged catchments in the Awash river basin using mean annual rainfall and catchment area. The coefficient of determination R^2 is 0.83 and the standard error of area and mean annual rainfall is 0.101 and 0.326 respectively, with 15 degrees of freedom (HALCROW, 1989).

Master plan study of a Rift valley lake basin (HALCROW G. L., 2007) regression analysis used to relate average annual flow (m^3/s) and catchment area (Km^2), longest drainage path (Km), slope, S_{1085}, altitude (m), annual average rainfall (mm), and annual average ETo (mm). Least-square stepwise regression analysis was performed and the catchment area was the only predictor that included in all model structures. Transforming the natural logarithm of the catchment characteristics data, only three predictors catchment area, average annual rainfall, and altitude were identified as appropriate models. Lastly, in view of the hydrological understanding of the basin, model fit assessment, and using R^2 and predictor R^2; catchment area and average annual rainfall was the best model that can be used to estimate average annual flow in ungauged catchments in the basin with R^2 of 0.72 and predicted R^2 of 0.65.

CHAPTER 3 METHODOLOGY

3.1 Description of Study Area

Abbay river basin is located in the northwestern region of Ethiopia between 7º 40' N and 12º 51' N latitude, and 34º 25' E and 39º 49' E longitude. It originates from the center of the catchment (Sekela, West Gojam), flows to the north into Lake Tana as Gilgel Abbay. Then, drain out from the southeast of Lake Tana, and flows through a deep gorge. Throughout the course, this river receives, Beshilo, Welaka, Jemma, Muger, Fincha, Guder, Didessa and Dabus from the east and south; and the Abeya, Suha, Chemoga, Birr, Fettam, Dura and Beles from the north and west. The Dinder and Rahad rise to the west of Lake Tana and flow westwards joining the Abbay river after crossing the border.

It's the second-largest basin next to the Wabisheble river basin which has a catchment area of (199,812 km^2), covering parts of Amhara, Oromia, and Benishangul-Gumuz regional states. It has the sixteen sub-basins of Anger, Beles, Beshilo, Dabus, Didesa, Dindir, Fincha, Guder, Jemma, Lake-Tana, Muger, North Gojam, Rahad, South Gojam, Welaka, Wonbera as shown in Figure 1.

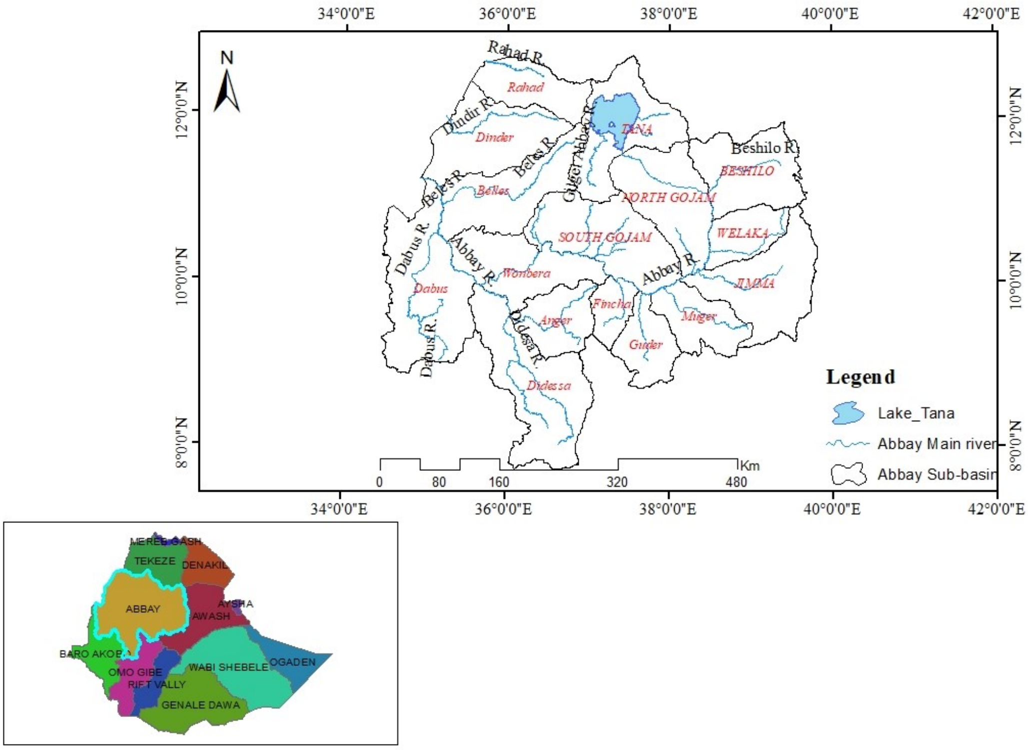


Figure 1 Study area

3.2 Data Availability and Material

3.2.1 General

The necessary data that are used for this thesis mainly are three. These are the meteorological data (Rainfall), the hydrological (Streamflow) and the catchment physiographic (Area, Slope, and Elevation).

3.2.2 Meteorological data

As it is known that rainfall data is one of the crucial input parameters for any model that estimates flow data. In this thesis rainfall data were gathered from the National Meteorological Agency (NMA). Due to the national security and sensitivity of recorded data, any required data of rain gauge stations will not be possible to get according to the authority rules. The study area of this thesis covers almost 20% of the country's land area, hence, 32 rainfall stations inside the basin were collected.

3.2.3 Hydrological data

Monthly flow record data of selected stations were collected from the Ministry of water, irrigation, and energy (MoWIE), Hydrology Department.

3.2.4 Physiographic data

Arc GIS 10.4 using the Digital Elevation Model (DEM, 30m) of the study area was used to estimate catchment physiographic by delineating the hydrometric stations as an outlet of watershed.

3.2.5 Materials

For this thesis the following materials used and their functions were summarized by the table below.

No.	Name	Function
1	Arc GIS 10.4	For delineation of the watershed, Estimation of physiographic data, and Estimation of areal rainfall
2	R Studio	For data splitting
3	Microsoft Excel	For Preparation of meteorological and hydrological data For the development of empirical equations

3.3 Method of Analysis

3.3.1 Hydrometric stations

Hydrometric stations were selected taking currently available designated stations in master plan studies of the basin and additional stations are added by considering spatial distribution on the study area. However, all required stations were not available in the directorate in addition to long periods of gaps in records.

Therefore, for this thesis hydrometric stations having complete year data and missing in lean flow seasons (November – May) of the year are considered.

Table 1 Selected hydrometric stations for Abbay river basin

No	Station No	River	Site	Lat_DD	Lon_DD
1	112001	Abbay	Nr. Kessi	11.07	38.18
2	112003	Abbay	@ Bahir Dar	11.6	37.4
3	112004	Andassa	Nr. Bahir Dar	11.5	37.48
4	114002	Angar	Nr. Nekemet	9.43	36.52

5	113029	Ardy	Nr. Metekel	10.95	36.52
6	112018	Azuari	Nr. Mota	10.97	38.02
7	113001	Bello	Nr. Guder	8.87	37.67
8	112007	Beressa	Nr. Debre Berhan	9.67	39.52
9	113013	Birr	Nr. Jiga	10.65	37.38
10	113008	Chemoga	Nr. Debre Markos	10.3	37.73
11	114001	Didessa	Nr. Arjo	8.68	36.42
12	113023	Dura	Nr. Metekel	10.98	36.48
13	111002	Gelgel Abbay	Nr. Maraw	11.37	37.03
14	113005	Guder	@ Guder	8.95	37.75
15	113012	Gudla	@ Dembech	10.55	37.5
16	111006	Gumara	Nr. Bahir Dar	11.83	37.63
17	115003	Hoha	Nr. Asoss	10.15	34.63
18	113011	Jedeb	Nr. Amanuel	10.4	37.57
19	111003	Koga	@ Merawi	11.37	37.05
20	113036	Lower Fettam	@ Galibed	10.48	37.02
21	111007	Megech	Nr. Azezo	12.48	37.45
22	112002	Mugher	Nr. Chancho	9.3	38.73
23	113026	Neshi	Nr. Shambu	9.75	37.25
24	112029	Robigumer	Nr. Lemi	9.75	39
25	112016	Shina	Nr. Adiet	11.25	37.5
26	113014	Temcha	Nr. Dembecha	10.53	37.5
27	112040	Wenka	Nr. Istay	11.62	38.07

Generally, 27 hydrometric stations are selected for this study;

Table 1 and Figure 2 summarized their location and spatial distributions in the basin. These stations are split into two datasets. The first one was for the development of equations or training datasets and the second one was for validation of developed equations or test datasets. The splitting processes were done in R studio based on 70% of the data for training and the remaining 30% for validation.

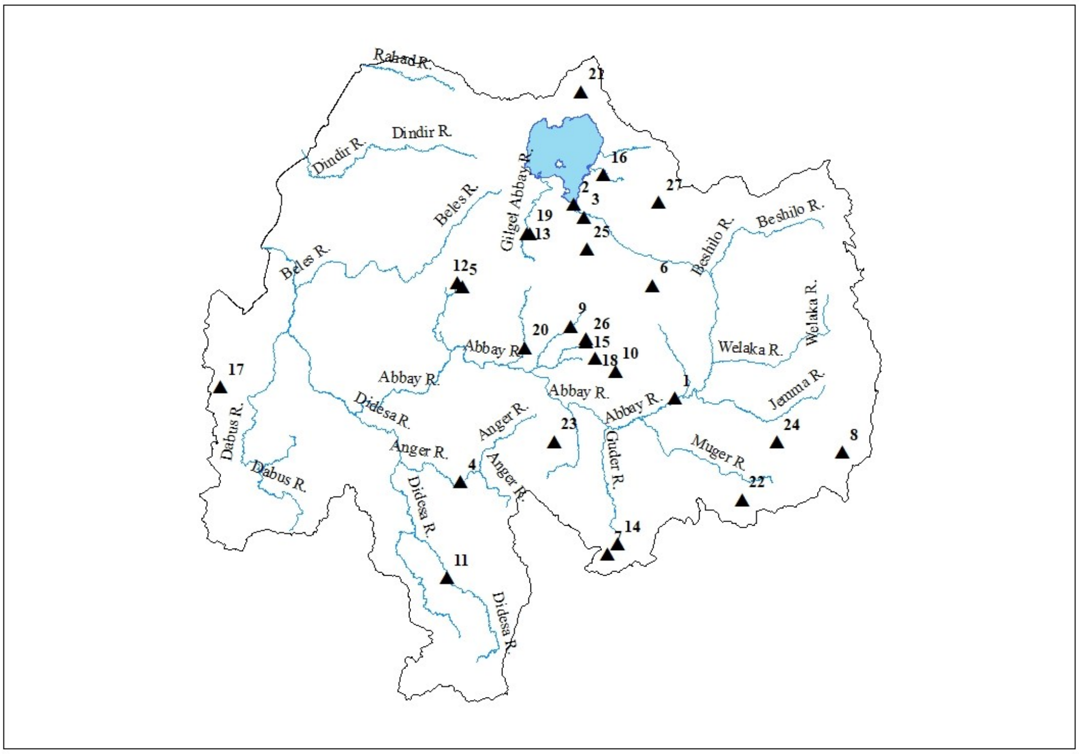


Figure 2 Abbay river basin selected hydrometric stations

3.3.2 Hydro-metrological data Analysis

In any water resource study different preliminary exploratory data analyses are constantly made. In this thesis, the following data analysis was done for the selected stations.

3.3.2.1 Meteorological data

Filling missing data: the first step in ground rainfall station analysis is filling the gap of missed records. In the Abbay basin, the possible available gauging stations which are recorded data for the period 1981 to 2013 are 32 in which spatially distributed in the basin except in the northwest side as shown in Figure 3.

Recorded data missed maybe because of a lack of appropriate records, shifting of the station location. This error should be given attention because they lead inconsistency and ambiguous results that may contradict to the actual situation.

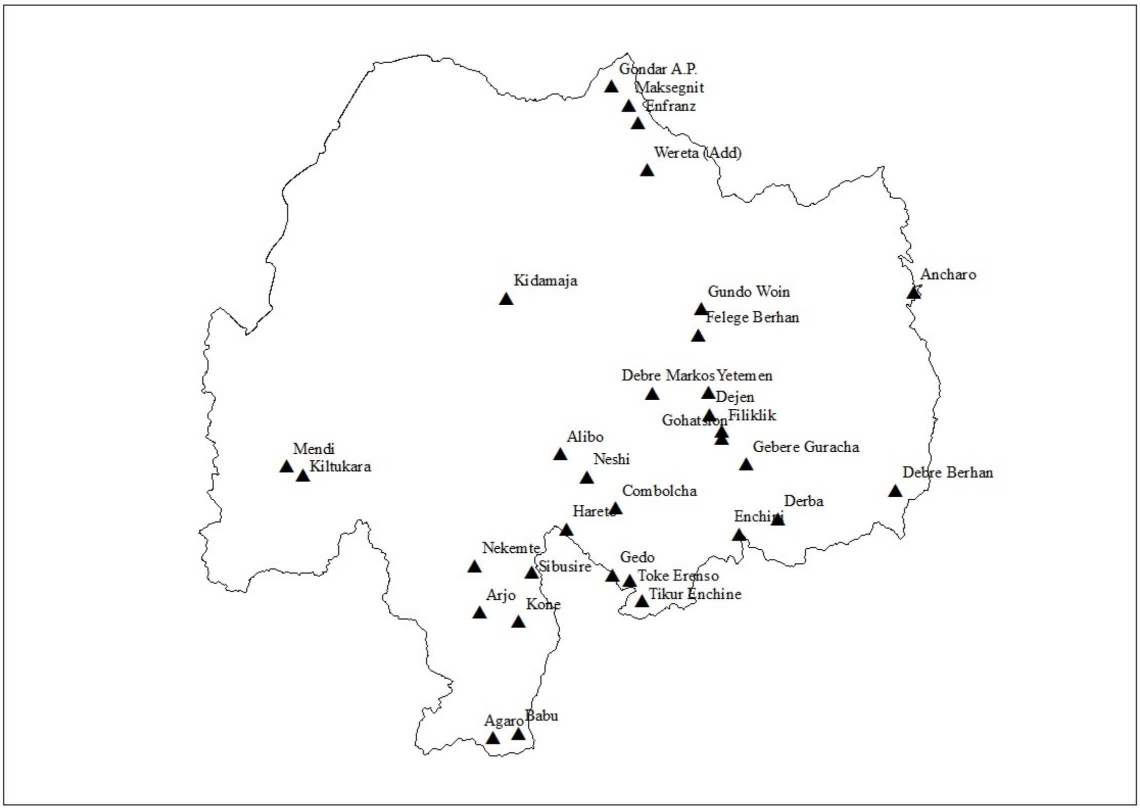


Figure 3 Rain gauge stations for Abbay Basin

There are different methods of tackle down this problem like Arithmetic, Normal ratio and linear regression method. For this thesis, the linear regression method of filling gaps was adopted. Stations (donor) that are close to the test stations are selected and their linear relationships are developed and there R^2 greater than 0.5 were adopted for filling. The order of the nearby stations is depending on their distance with the test station.

Table 2 Debre Berhan rain gauge station priority and regression equation used for infilling

Test station	Missed data	Regression Eq.	Rain gauge Priority for infilling			
		Pt = α*Pd	1	2	3	4
Debre Berhan	4.80%	Donor gauge	Kotu	Gudoberet	Deneba	Debre Sina
		α	1.09	0.832	0.647	0.526
		R2	0.93	0.72	0.77	0.66

In case if the test and primary donor stations have the same missed data the second-order donor station was used until the gap in test station fill. For instance, Table 2 summarizes the Debre Berhan rain gauge station filling of missed data analysis. Moreover, Figure 4 also shows the developed regression equation and R^2 of Debre Berhan with Kotu rain gauge station. In the same fashion, the rest of the station's analysis can be found in Appendix-A.

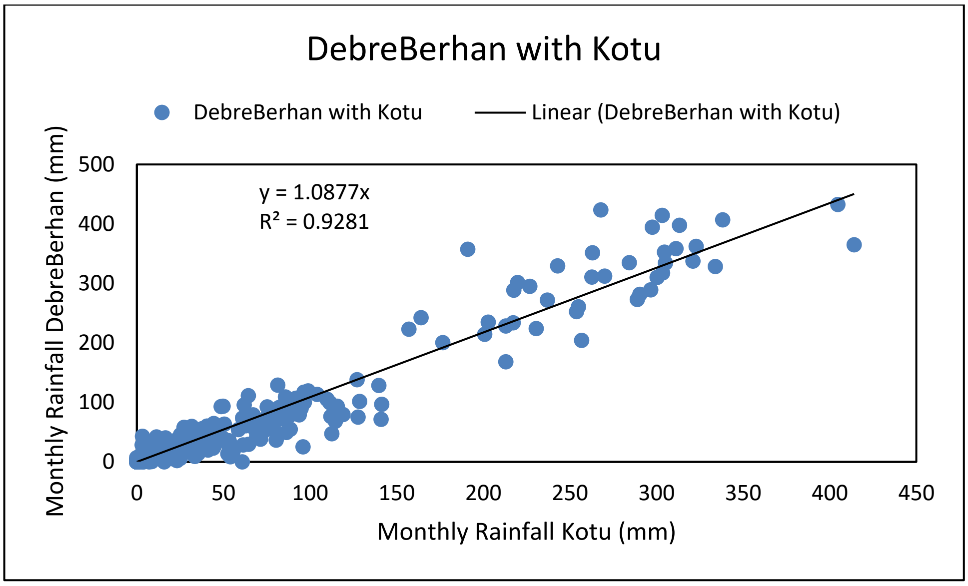


Figure 4 Regression analysis for Debre Berhan using Kotu as a donor rain gauge

Homogeneity test: Inhomogeneity's in station data records are often caused by changes in observational routines, among which are station relocations, changes in measuring techniques and changes in observing practices (Wijngaard et al., 2003). The homogeneity of annual precipitation of each station is tested using XLSTAT in which incorporates four homogeneity tests namely Pettitt, Buishand, SNHT, and von Neumann. The tests presented in this tool correspond to the alternative hypothesis of a single shift. For all tests, XLSTAT provides p-values using Monte Carlo resampling.

Hence, the hypothesis for this test is

H_0: Data are homogeneous

H_a: There is a year at which there is a change in the data

As the computed p-value is greater than the significance level alpha=0.05, the null hypothesis H_0 accepted.

After the analysis of homogeneity, the test results are classified as follows (Wijngaard et al., 2003)

Class A: Useful

The series that rejects one or none null hypothesis under the four tests at a 5% significance level are considered. Under this class, the series is grouped as homogeneous and can be used for further analysis. For example, the Combolcha rain gauge station falls under this class in which its test value Table 3 and graphical result Figure 5 are shown below.

Table 3 Homogeneity test for Combolcha rain gauge station

Station/Test	Pettitt	SNHT test	Buishand	von Neumann
Combolcha	0.991	0.706	0.406	0.877

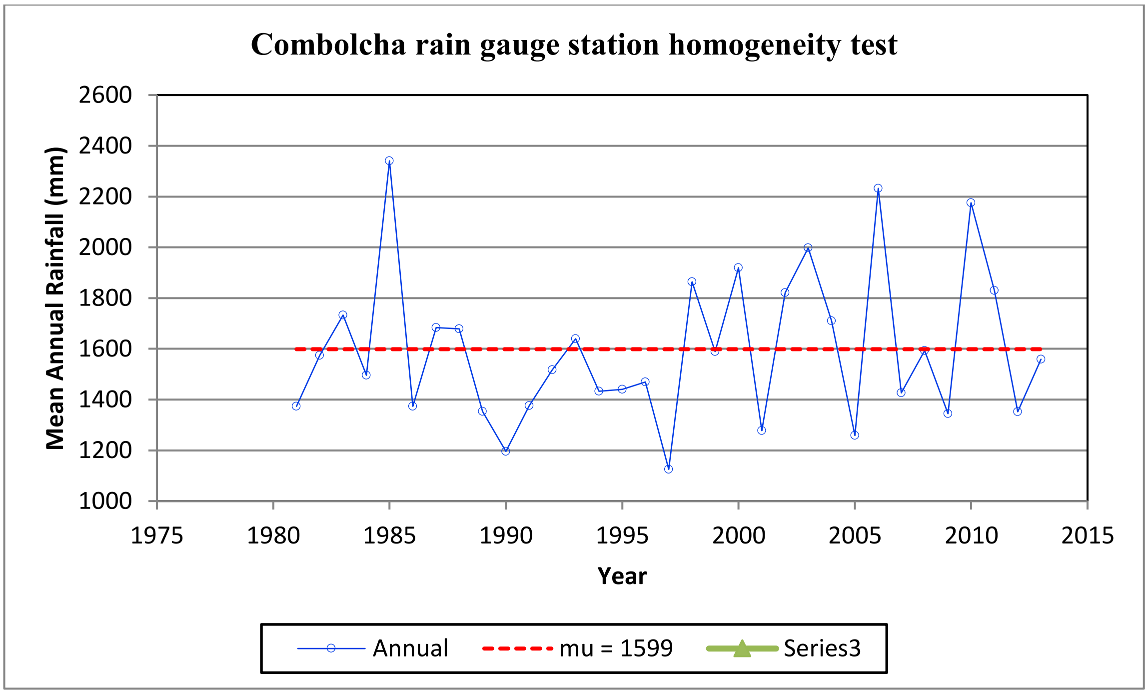


Figure 5 Homogeneity test for Combolcha rain gauge station

Class B: Doubtful

The series that reject two null hypotheses of the four tests at a 5% significance level is placed in this class. In this class, the series has an inhomogeneous signal and should be critically inspected before further analysis. For this analysis the only station which falls under this class is Kone. If we look at the result (Table 4 and Figure 6) Buishand and SNHT tests are the failures however, it’s possible to say Buishand test is fulfilled and considering trend test which can be referred to in the next section that shows there is no trend. Hence, this station is included for further analysis.

Table 4 Homogeneity test for Kone rain gauge station

Station/Test	Pettitt	SNHT test	Buishand	von Neumann
Kone	0.285	**0.041**	**0.047**	0.167

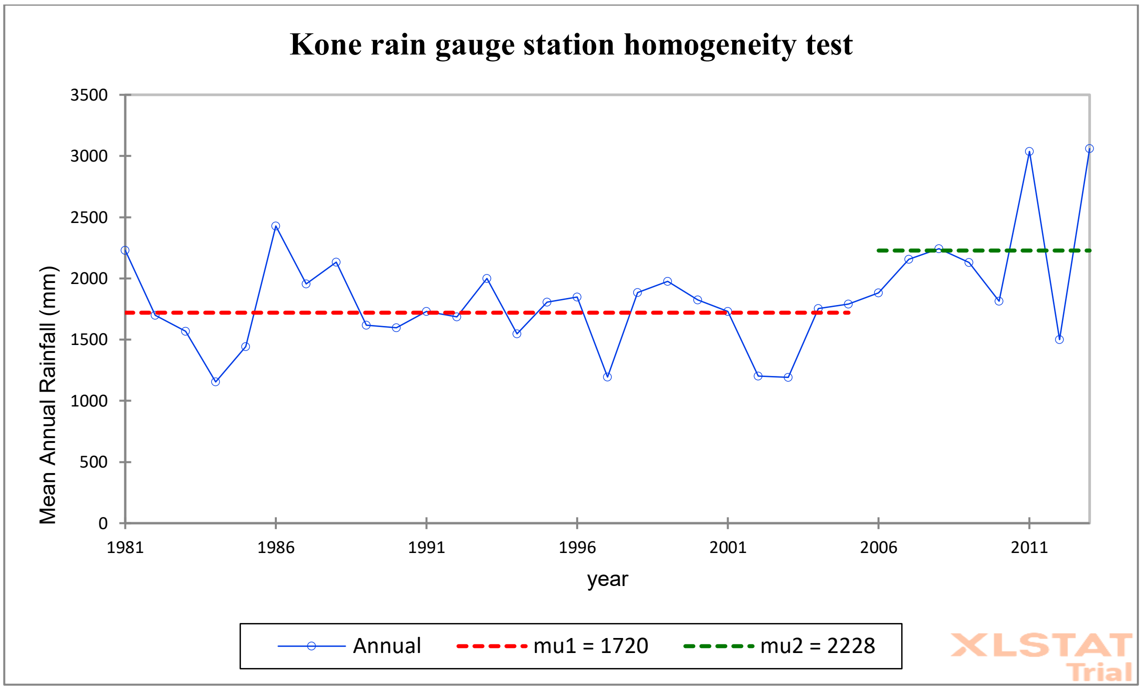


Figure 6 Homogeneity test for Kone rain gauge station

Class C: Suspect

When there are three or all tests are rejecting the null hypothesis at a 5% significance level, then the series is classified into this category. In this category, the series can be deleted or ignored before further analysis. For instance, the station which is under this class is summarized in Table 5 and Figure 7 as follows.

Table 5 Homogeneity test for Toke Erneso rain gauge station

Station/Test	Pettitt	SNHT test	Buishand	von Neumann
Toke Erneso	**0.000**	**< 0.0001**	**< 0.0001**	**0.000**

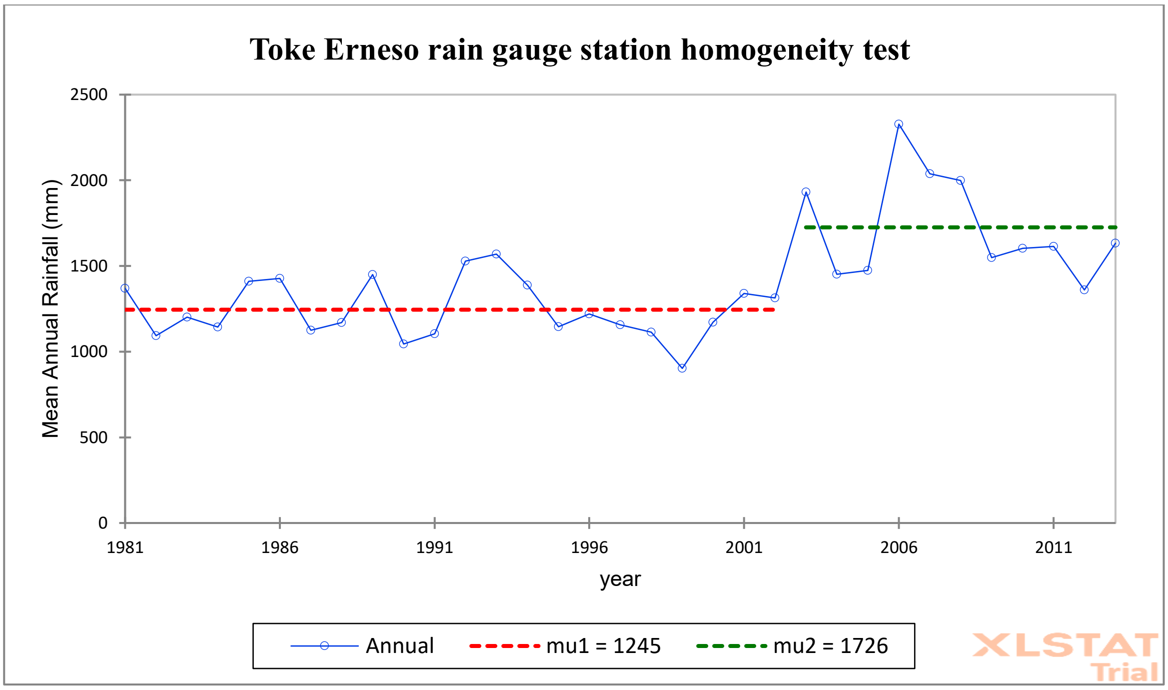


Figure 7 Homogeneity test for Toke Erneso rain gauge station

Generally, from the total 32 rain gauge stations class A useful 21 (65.6%), class B doubtful 1 (3.1%) and class C suspect 10 (31.3%). However, doubtful and suspect stations can be checked and interpret using the history of the station location, observing practice and instruments used. Unluckily, this information is not available hence, class C stations are excluded from further analysis. The whole results of the homogeneity test found in Appendix – C.

Trend test: Trend analysis also checked by XLSTAT using the non-parametric Mann Kendall (MK) trend test which is well known and mostly adopted. The hypothesis in this test is that

H_0: There is no trend in the series

H_a: There is a trend in the series

As the computed p-value is lower than the significance level alpha=0.05, one should reject the null hypothesis H_0, and accept the alternative hypothesis H_a.

Additionally, using Sen's slope magnitude it is possible to conclude the trend type is increasing as well as decreasing trend. Hence, if the Sen's slope is positive it is increasing and if it is negative it is decreasing.

Therefore, based on the above hypothesis p-value and Sen's slope trend analysis summarized in the tabular form below.

Table 6 Trend analysis for available rain gauge stations

Station Name	Mann Kendall Test			Sen's Slope	Test Interpretation	Trend type
	Kendall's tau	p-value	S			
Agaro	-0.05	0.70	-26	-2.88	Accept H0	Decreasing
Alibo	0.05	0.72	24	1.25	Accept H0	Increasing
Ancharo	0.00	0.99	-2	-0.10	Accept H0	Decreasing
Arjo	0.42	0.00	222	21.17	*Reject H0*	Increasing
Babu	-0.06	0.65	-30	-2.91	Accept H0	Decreasing
Combolcha	0.07	0.59	36	4.15	Accept H0	Increasing
DebreBerhan	0.34	0.01	178	6.25	*Reject H0*	Increasing
DebreMarkos	0.13	0.30	68	3.22	Accept H0	Increasing
Dejen	0.08	0.53	42	2.74	Accept H0	Increasing
Derba	-0.28	0.03	-146	-12.37	*Reject H0*	Increasing
Enchini	0.00	0.99	-2	-0.16	Accept H0	Decreasing
Enfranze	0.07	0.57	38	1.71	Accept H0	Increasing
Felege Berhan	0.02	0.89	10	1.07	Accept H0	Increasing
Filkilik	0.13	0.30	68	7.25	Accept H0	Increasing
Gebre Guracha	0.16	0.19	86	5.03	Accept H0	Increasing
Gedo	-0.23	0.07	-120	-8.52	Accept H0	Decreasing
Gohatsion	0.17	0.17	90	3.82	Accept H0	Increasing
Gonder A.P.	0.28	0.03	146	6.39	*Reject H0*	Increasing
Gundo Woin	0.28	0.02	150	7.67	*Reject H0*	Increasing
Hareto	-0.11	0.38	-58	-5.64	Accept H0	Decreasing
Kidamaja	0.08	0.51	44	2.76	Accept H0	Increasing
Kiltukara	0.51	< 0.0001	268	57.38	*Reject H0*	Increasing
Kone	0.19	0.13	100	10.72	Accept H0	Increasing
Makisegnet	0.06	0.63	32	1.93	Accept H0	Increasing
Mendi	0.23	0.06	122	14.06	Accept H0	Increasing
Nekemt	0.15	0.22	80	6.22	Accept H0	Increasing
Neshi	0.33	0.01	176	25.31	*Reject H0*	Increasing

Sibusire	-0.05	0.68	-28	-2.08	Accept H0	Decreasing
Tikure Enchine	0.33	0.17	174	13.55	Accept H1	Increasing
Toke Erneso	0.40	0.00	210	15.35	*Reject H0*	Increasing
Woreta	0.03	0.84	14	0.75	Accept H0	Increasing
Yetmen	0.36	0.30	192	13.95	Accept H1	Increasing

As can be deduced from Table 6 the only 8 stations (Bold) are unsuccessful to satisfy the hypothesis as no trend. So, considering the above two tests 22 stations were used for this analysis.

Monthly Rainfall Data: this analysis started with calculating an arithmetic average of each month of 22-gauge stations with record period 1981-2013 and taking it as an estimated basin-wide mean monthly rainfall.

Table 7 Abbay basin mean monthly rainfall statistics

Month	Mean (mm)	Standard deviation (mm)	C.v.
January	13.35	9.19	0.69
February	16.53	10.26	0.62
March	54.13	23.62	0.44
April	69.70	26.00	0.37
May	125.39	55.62	0.44
Jun	202.39	82.46	0.41
July	322.43	63.39	0.20
August	313.81	67.60	0.22
September	190.24	66.64	0.35
October	78.67	39.75	0.51
November	28.42	12.64	0.44
December	15.20	8.54	0.56

Where C.v. coefficient of variation (Mean/Standard deviation).

As it is seen from Table 7 and Figure 8 wet months of the basin are June, July, August, and September and the dry season months are January, February, March, April, November, and December. May and October are the transition month from the dry season to the wet season and the transition month from the wet season to the dry season respectively. January has the lowest (13.35mm) and July (322.43mm) has the highest monthly mean rainfall.

The coefficient of variation, C.v. of the wet season months is small indicating that these months have low variation. The dry season months have a higher coefficient of variation indicating that the year-to-year variation for these months is high

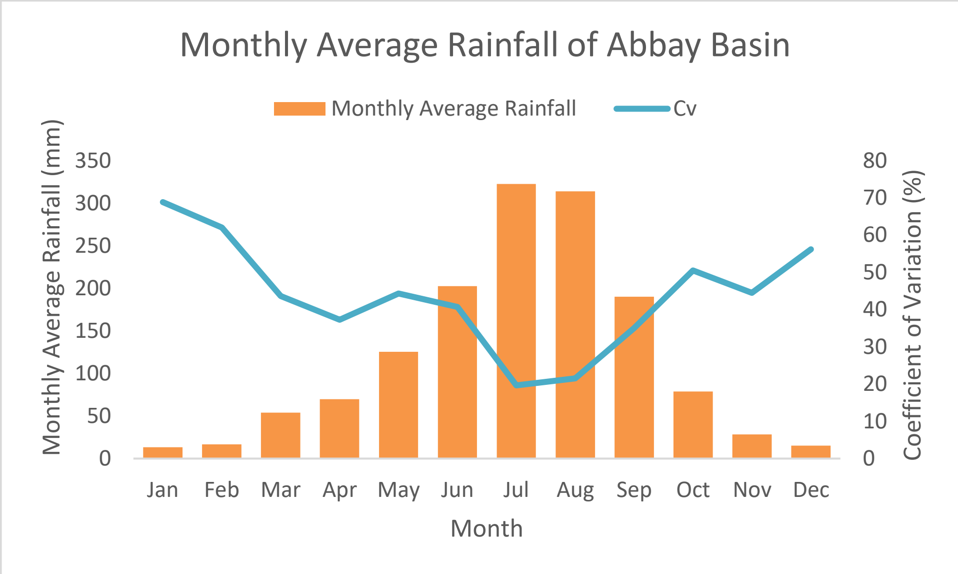


Figure 8 Mean monthly rainfall of the Abbay River Basin (1981-2013)

Annual Rainfall Data: The mean annual rainfall of the Abbay river basin is computed as the arithmetic means of the annual rainfall for the year under consideration of each station. So, the average annual rainfall of the year under consideration computed from 22 gauges. The spatial distribution of mean annual rainfall for the Abbay basin is shown in Figure 9 below. The mean annual rainfall using 22 gauges all over the basin is 1430mm.

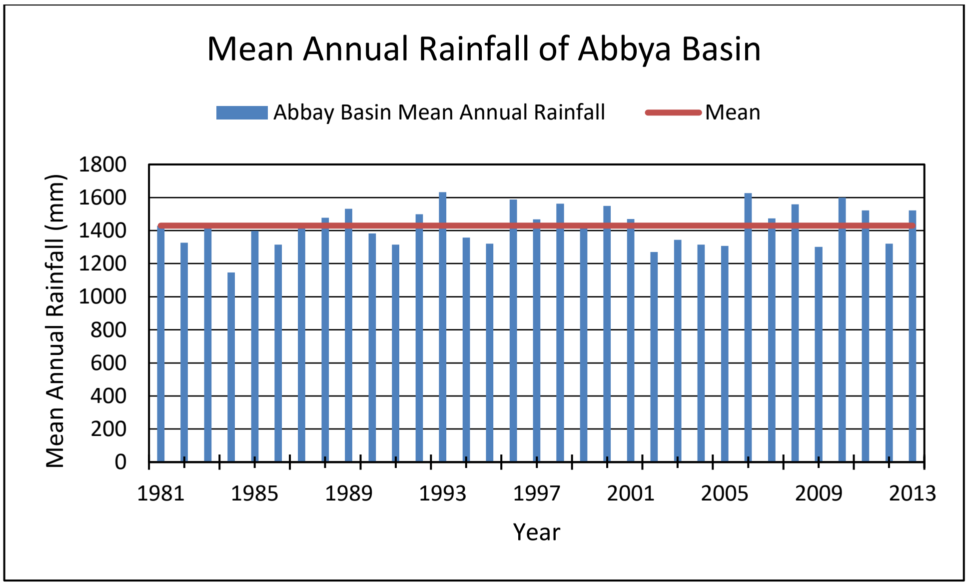


Figure 9 Abbay river basin mean annual rainfall (1981-2013)

Areal Rainfall: Various interpolation methods have been developed for this purpose, generally deterministic and geostatistical interpolation methods. The first one is simple techniques which are based on the location of the measuring stations and on measured values; For instance, the most commonly used Thiessen polygons method and inverse distance weighting method are examples of deterministic methods of interpolation. The later method of interpolation is based on statistical models that include autocorrelation which is statistical relationships among the measured points. In this thesis, the Kriging method of interpolation is selected. Kriging is similar to inverse distance weighting (IDW) in that it weights the surrounding measured values to derive a prediction for each location. However, the weights are based not only on the distance between the measured points and the prediction location but also on the overall spatial arrangement among the measured points.

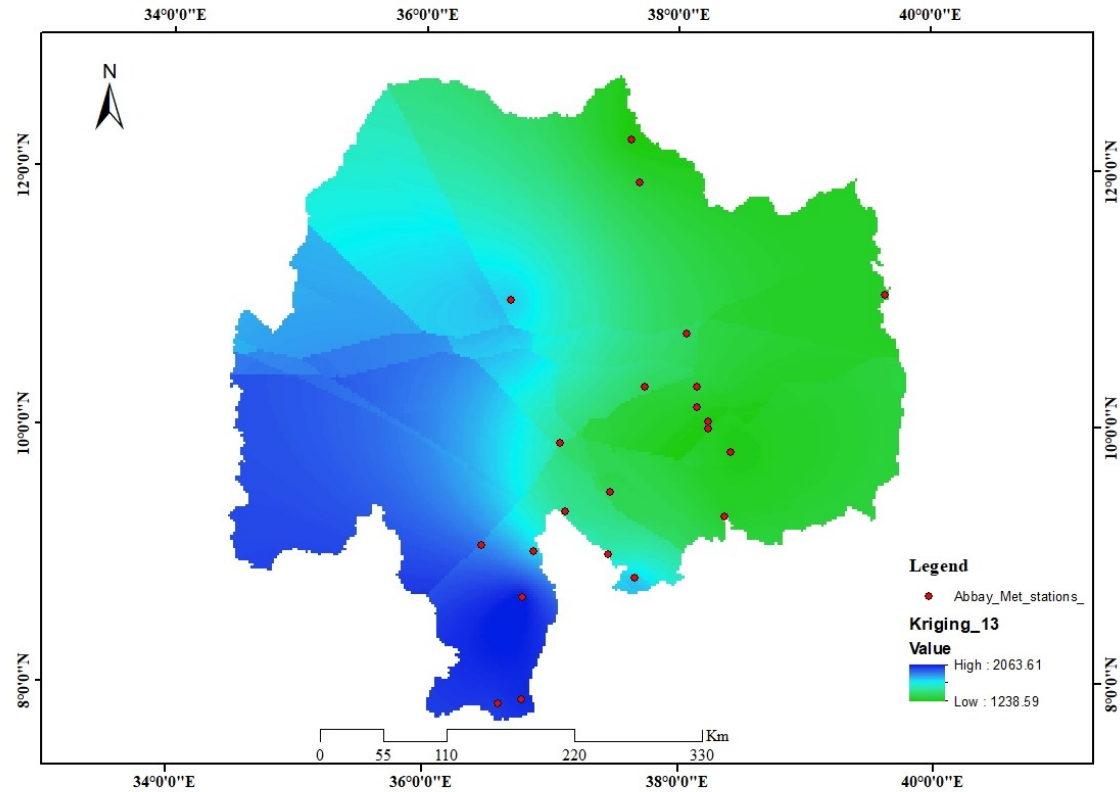


Figure 10 Areal rainfall interpolation using Kriging for the year 2013

Ordinary kriging method of interpolation was established for each consecutive year from 1981-2013 with 22 rain gauges using Arc GIS 10.4 spatial analysis tool. As a sample Figure 10, exhibits the interpolation raster layer for record period 2013 with the location of rain gauge stations used. Additionally, the delineated area of each hydrometric station used to extract the annual rainfall.

3.3.2.2 Hydrometric Stations

Filling missing data: The streamflow data obtained from MoWIE, HD for the selected station have missed data in addition to a smaller number of records. In this thesis gap-filling only conducted for stations that had missing in the lean flow seasons (November – May) by taking all the available data of the considered month. For example, Abbay near Kessie station the available record period from 1956-2013 was collected from MoWIE but for this study record period from 1981 – 2013 was selected. From these selected periods there are missed record months. However, only 2 record years 2002 and 2008 are missed data in lean flow months December and February respectively. So, the gap was filled by taking an average of the same month from the total recorded year of the station.

Trend detection: the detection of the trend was done by plotting recorded year with annual flow and observations of unsupported trend patterns; All of the stations selected for analysis were exhibit supported trend. For example, RobiGumero near Lemi gauge station exhibits a good trend in record years.

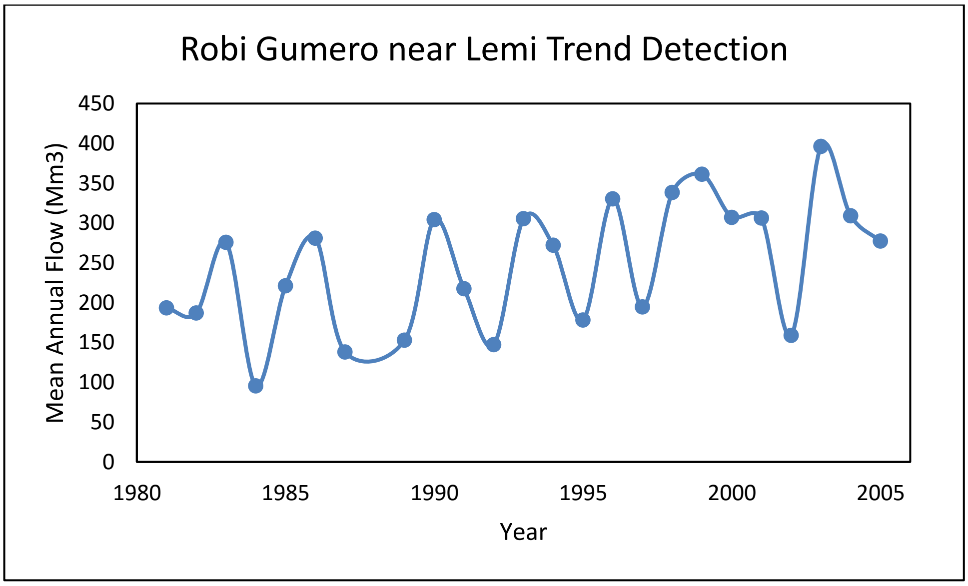


Figure 11 RobiGumero near Lemi trend detection

Outlier test: due to numerous reasons like observation error, equipment error and exceptional cases streamflow records may show significant change and this will affect the analysis of the data. From the different methods of outlier detection, the boxplot method which is the most adopted and show position, dispersion and skewness of data are used for this research. Box plot method of outlier detection done using the first quartile (Q1), third quartile (Q3), and inter Quartile (IQ).

Lower limit = Q1 – 1.5IQ and Upper Limit = Q3 + 1.5IQ

Where IQ = Q3 – Q1

As an example, Figure 12 and Table 8 show the outlier test for Megech near Azezo gauging station.

Table 8 Outlier test for Megech near Azezo gauging station

Minimum	71.32	Mean	279.37
First Quartile (Q1)	134.40	Inter Quartile range (IQR = Q3-Q1)	154.46
Median	199.12	IQR *1.5	231.69
Third Quartile (Q3)	288.87	Lower Limit	-97.29
Maximum	813.59	Upper Limit	520.56

A result from the box plot summarized in the above Table 8 shows the lower limit is negative which indicates there is no outlier, but the upper limit is small compared with the maximum value of the station. Hence, as it is seen from Figure 12 the next maximum value of the upper limit is 391.445 and records year which are labeled in a dot are outliers that have record greater than the upper limit.

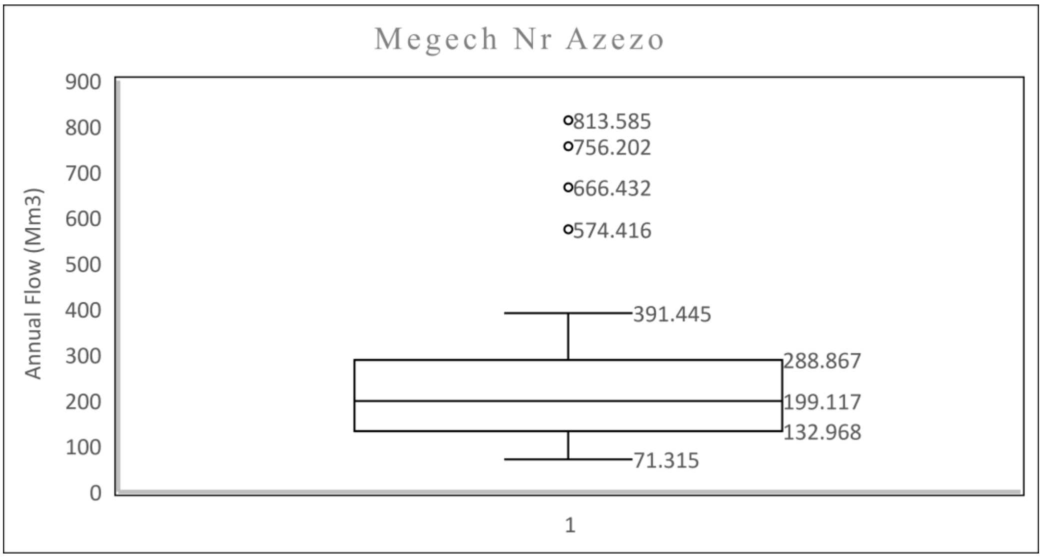


Figure 12 Outlier test using boxplot for Megech near Azezo station

However, as stated the outlier may be due to different reasons, hence, concluding by only considering the above equation will lead to a mistake. Therefore, comparing the flow data with rainfall data and also checking which month of flow will the gauge depended on is checked. So, when it checks the outlined outlier record years, they are starting from 2008 up to 2013. Hence, the monthly streamflow record data starting from 2007 exhibit radical change of values for each month and the rainfall record is not high. Therefore, record data outlined as outliers are eliminated from further analysis.

Additionally, there was an effort to develop a relationship between run off with rainfall in such case some data records which disturb the relationship between runoff and rainfall. The disturber record data is identified from the annual flow (Q) with the annual rainfall (P) plot of each selected station. As it is seen from Figure 13 for Birr near Jiga station left top two record years (high flow with low rainfall) and the right bottom one record year (low flow with high rainfall) disturb the relationship between rainfall and runoff.

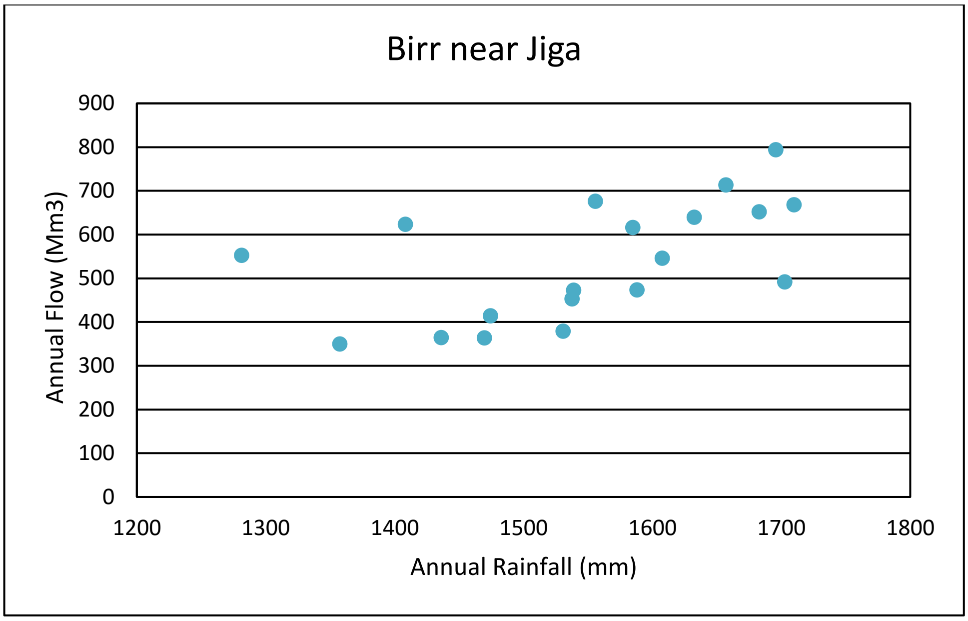


Figure 13 Rainfall-Runoff relationship of Birr near Jiga station with disturber

These types of disturbers were further checked if they were in dry periods. For Figure 13 scatter plot mentioned disturber are not in drought years so, they were eliminated from the analysis. The relationship without the disturber looks like Figure 14.

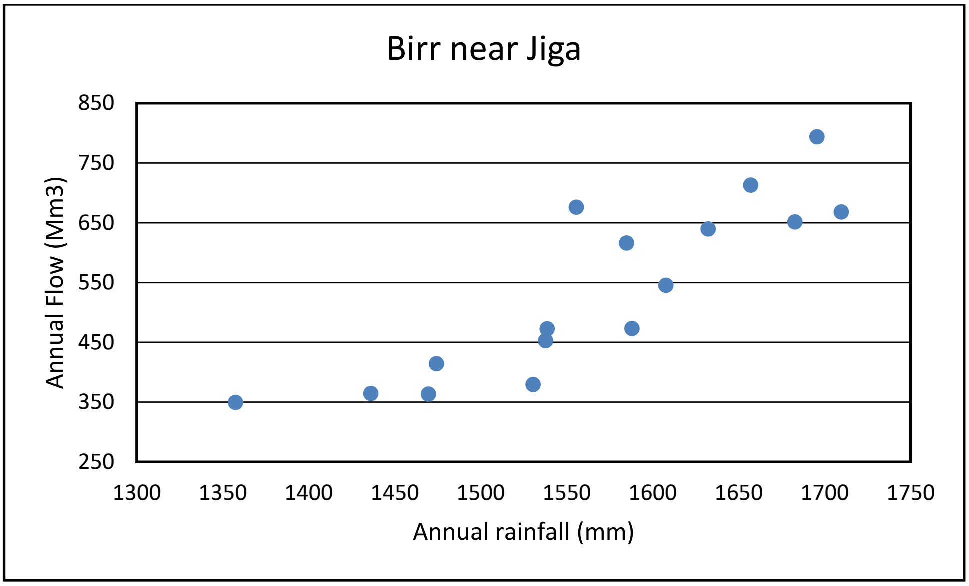


Figure 14 Rainfall-Runoff relationship of Birr near Jiga station without disturber

3.3.3 Temporal and Seasonal variability of Flow

According to Belete B et al. (2014), the surface water temporal variability follows similar rainfall variability patterns and seasonal variability of flow also depends on the number of rainfall season (bimodal and unimodal). Therefore, for this thesis, a few selected hydrometric station's temporal and seasonal variability of flow were performed as follows.

Abbay near Kessie

This station is the major gauging station placed in the main stem of the basin having a total area of 65,784 km^2. It is located downstream of the Addis Ababa - Bihar Dar road bridge that crosses the main stem of the Abbay river basin.

Time series of monthly flow data at this station for the period 1956-2013, obtained from MoWIE used for analysis. July, August, September, and October months exhibit high discharge over 58 years of records with a mean monthly flow of 3247 Mm^3, 6678 Mm^3, 3096 Mm^3, and 1434 Mm^3 respectively (Figure 16); with the highest peak in August of 13582 Mm^3 (Figure 15). The mean annual flow of 17.84 Bm^3 is observed at the station.

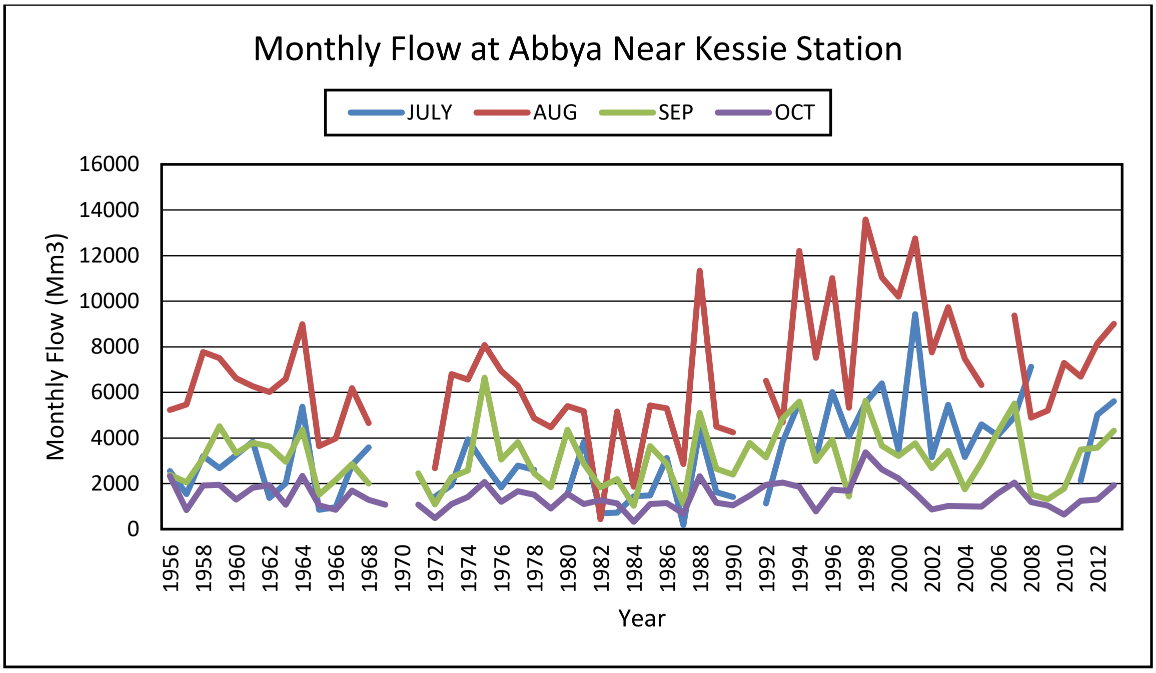


Figure 15 Monthly flow at Abbay near Kessie Station

In this station 6 record years (1981, 1985, 1986, 1991, 2009 and 2010) were not used; as stated in section 3.3.2.2 record period having missed data in good flow season (June to September) was eliminated from the analysis.

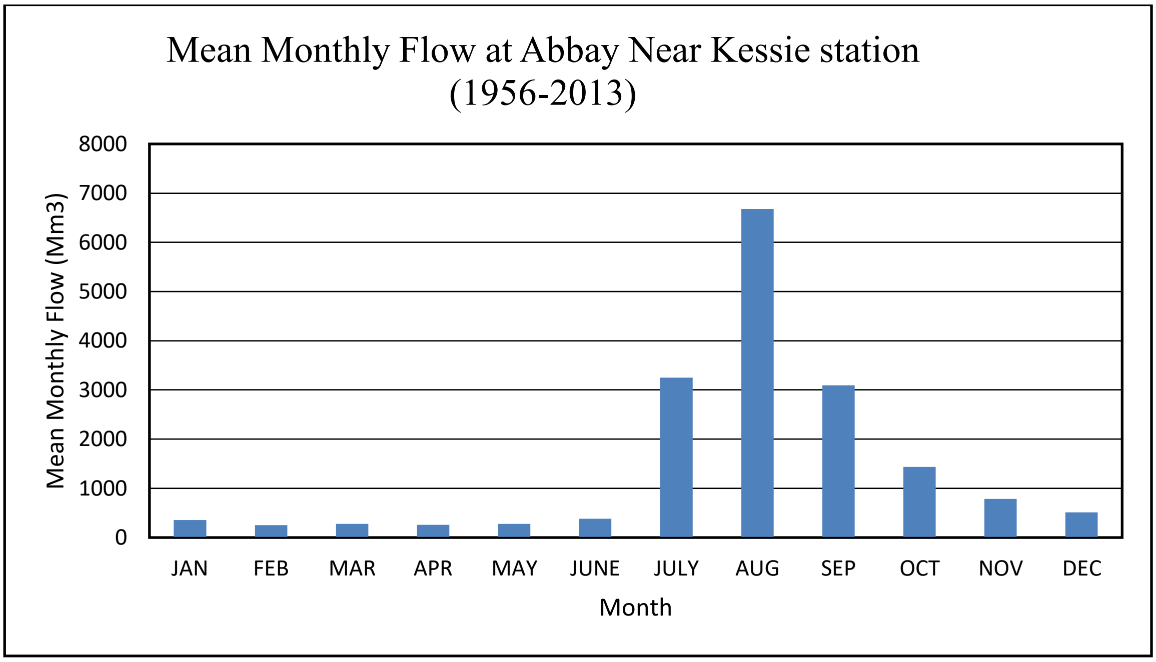


Figure 16 Mean monthly flow at Abbay near Kessie station

Beressa near Debre Berhan Station

This station located at the tip of the east of the basin near Debre Berhan town with a drainage area of 211km^2.

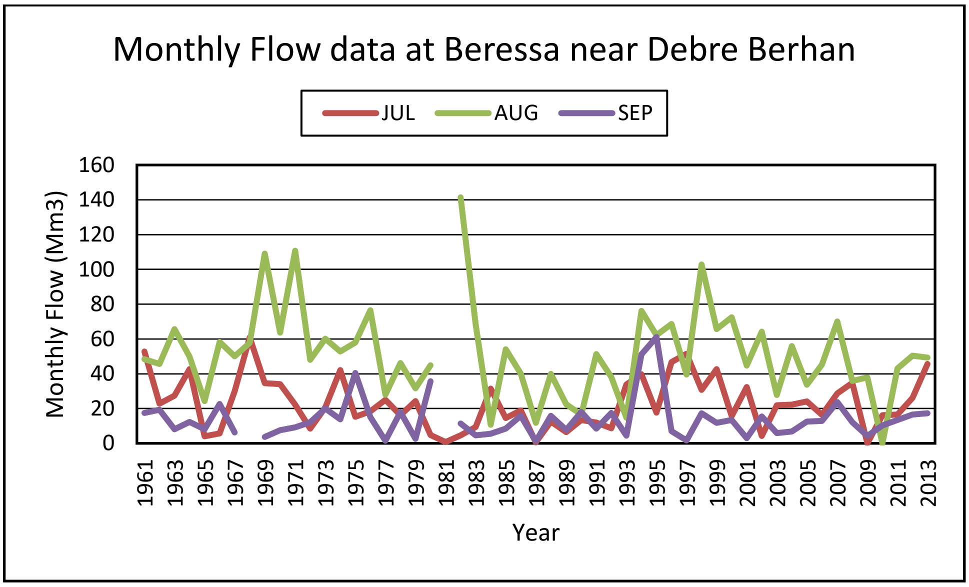


Figure 17 Monthly flow at Beressa near Debre Berhan Station

Long record period 1961-2013 were collected from MoWIE. Figure 17 exhibits that the dependable flow month of the station in July, August, and September with 23Mm^3, 53Mm^{3,} and 14Mm^3 respectively. Figure 18 shows that August is the peak month of 141Mm^3. Additionally, a mean annual flow of 97Mm^3 is observed. The recording period that used in this thesis record year 1981, 2009 and 2010 was not included for analysis with a noticeably missed month in good flow seasons.

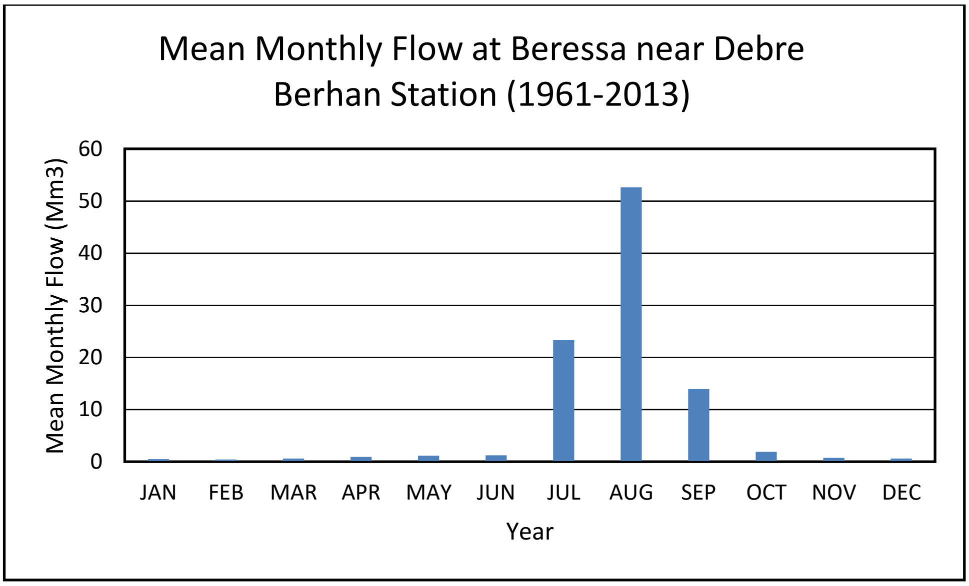


Figure 18 Mean monthly flow at Beressa near Debre Berhan station

Gumara near Bahir Dar Station

Gumara near Bahir Dar Station is located in the east of Bahir Dar to Gonder road near to the bridge which has a drainage area of 1,394 km^2. Gumara watershed is one of the major tributaries of Lake Tana.

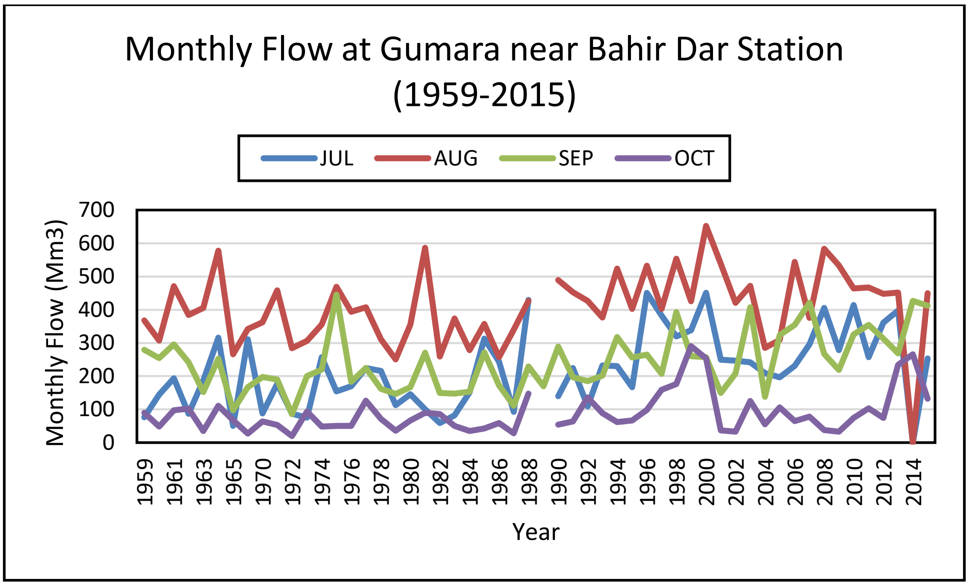


Figure 19 Monthly Flow at Gumara near Bahir Dar Station

Record period from 1959 to 2015 with missed the record year 1967 – 1969 collected from MoWIE. As Figure 20 shows July, August, September, and October are the highest flow months with 223Mm^3, 414Mm^3, 242Mm^3, and 89Mm^3 respectively; peak flow month in

August recorded 652Mm^3 (Figure 19). The mean annual flow of the station 1062Mm^3 observed.

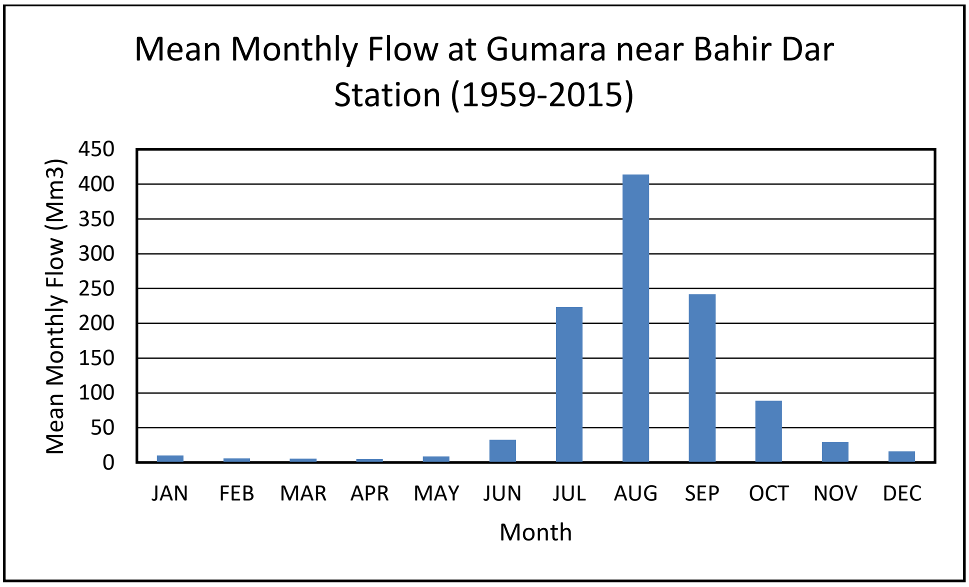


Figure 20 Mean monthly flow at Gumara near Bahir Dar station

From the base study period, 1981 to 2013 the record year 1989 was eliminated from the analysis due to missed month in good flow season (July and August).

Koga near Merawi Station

The Koga river is a major tributary of Gilgel Abbay that flows into Lake Tana. Koga hydrometric station located near Merawi town with a drainage area of 244Km^2. Monthly flow data for record period 1959-2012 collected from MoWIE. From Figure 22 July, August, September, and October are the highest flow months with 28Mm^3, 47Mm^3, 31Mm^3, and 16Mm3 respectively; peak flow record in August 254Mm^3 (Figure 21). The mean annual flow of 158Mm^3 observed in the station.

In this station record year 1968, 1969 and 1980 some months and all yearly records of 1981 and 1982 were estimated, hence, it did not include in the above calculations.

Additionally, for the study period the record year, 1990 and 1991 excluded from the analysis with the reason mentioned in previous stations.

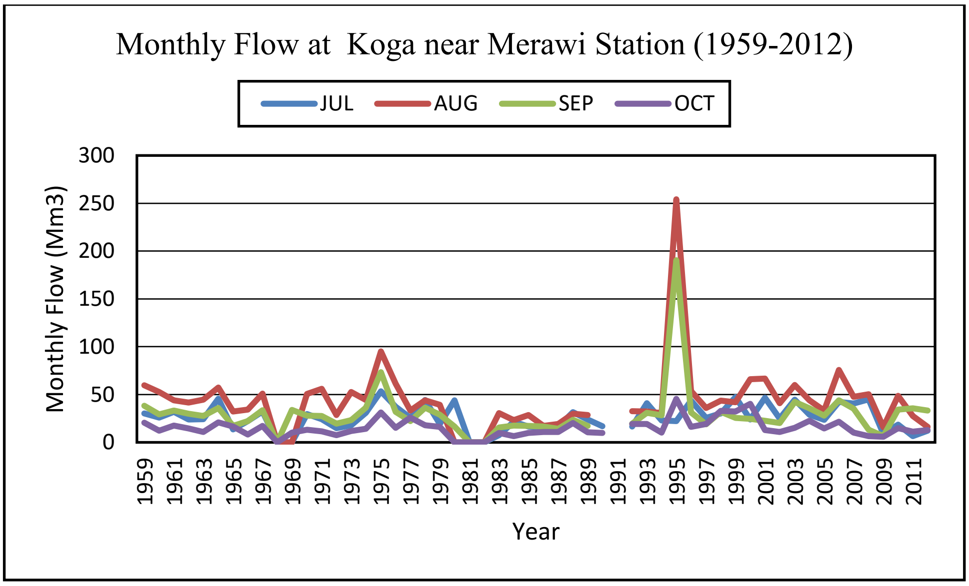


Figure 21 Monthly Flow at Koga near Merawi station

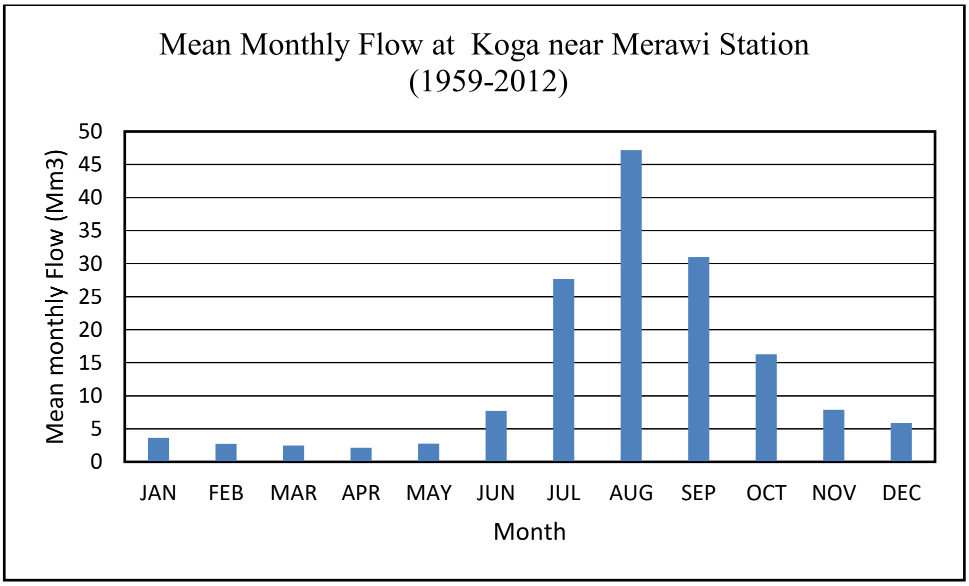


Figure 22 Mean monthly flow at Koga near Merawi station

Megech near Azezo Station

Megech river is one of the major tributaries of Lake Tana in which flow from the north. The hydrometric station located near Azezo town that records for the period 1956 to 2014 which the data obtained from MoWIE, the drainage area of 462 Km^2. The record year 1969, 1970, 1975, 1976, and 1979 were not included. Figure 24 exhibits that July, August, and September are the highest flow months with 48Mm^3, 110Mm^3 and 40Mm^3 respectively; 110Mm^3 of peak flow in the month August (Figure 23).

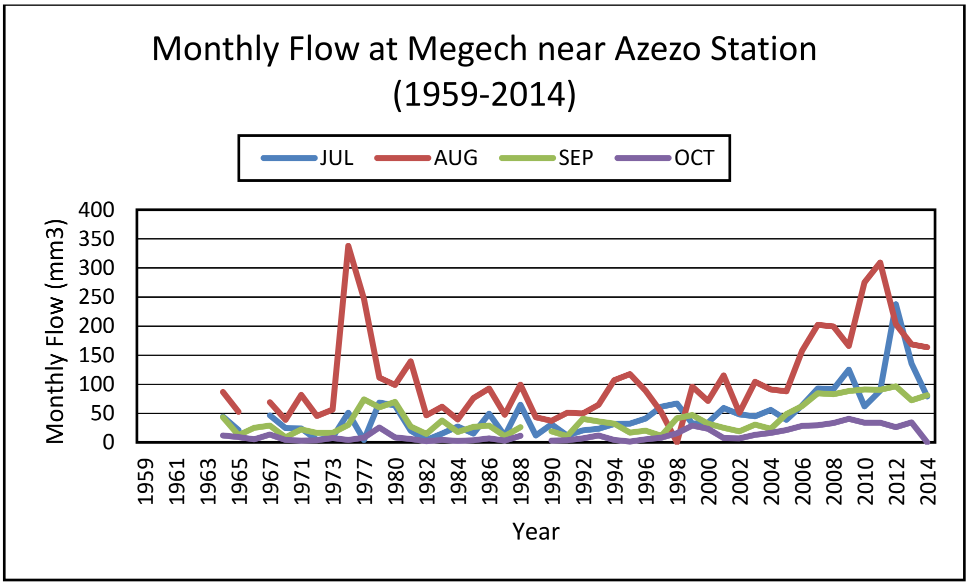


Figure 23 Monthly flow at Megech near Azezo station

The mean annual flow of this hydrometric station is 257Mm^3. Due to missing in July, August and September months record years 1989, 1998 and 2007 were excluded from further analysis in this study.

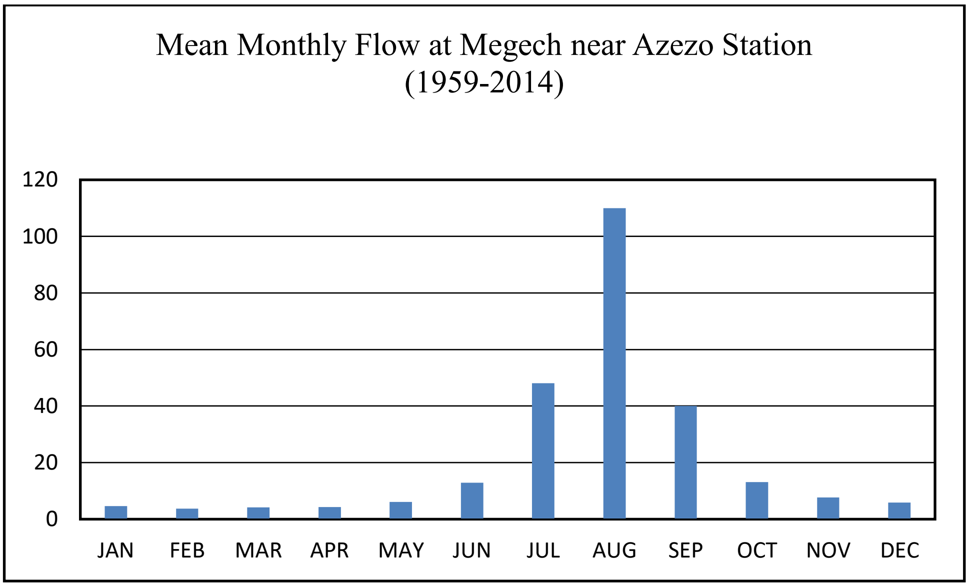


Figure 24 Mean monthly flow at Megech near Azezo station

3.4 Physiographic Data Analysis

3.4.1 Delineation of watersheds

Watersheds, also known as basins or catchments, are physically delineated by the area upstream from a specified outlet point. Watersheds can be delineated manually using paper maps, or digitally in a GIS environment.

The watershed delineation process was done using Arc GIS 10.4 extension Arc hydro tool. Using a 30m digital elevation model (DEM) as input, and delineation steps Fill sinks, Flow direction, Flow accumulation, Stream definition, Stream segmentation, Catchment grid delineation, Catchment polygon processing, Drainage line processing, and Adjoint catchment processing were done respectively. Lastly, using point delineation and each hydrometric station location as an outlet the delineation is completed.

Table 9 Comparison of a delineated area with hydrology directorate area

No	Sta. No	River	Site	Lat_DD	Lon_DD	HD Area	Delineated Area
1	112001	Abbay	Nr. Kessi	11.07	38.18	65784	64259
2	112003	Abbay	@ Bahir Dar	11.6	37.4	15321	15090
3	112004	Andassa	Nr. Bahir Dar	11.5	37.48	573	586
4	114002	Angar	Nr. Nekem	9.43	36.52	4674	4641
5	113029	Ardy	Nr. Metek	10.95	36.52	219	259
6	112018	Azuari	Nr. Mota	10.97	38.02	209	203
7	113001	Bello	Nr. Guder	8.87	37.67	290	292
8	112007	Beressa	Nr. Debreber	9.67	39.52	211	205
9	113013	Birr	Nr. Jiga	10.65	37.38	978	978
10	113008	Chemoga	Nr. Debrema	10.3	37.73	364	360
11	114001	Didessa	Nr. Arjo	8.68	36.42	9981	9974
12	113023	Dura	Nr. Metek	10.98	36.48	539	541
13	111002	Gel.Abbay	Nr. Maraw	11.37	37.03	1664	1671
14	113005	Guder	@ Guder	8.95	37.75	524	592
15	113012	Gudla	@ Dembech	10.55	37.5	242	245
16	111006	Gumara	Nr. Bahir Dar	11.83	37.63	1394	1355
17	115003	Hoha	Nr. Asoss	10.15	34.63	161	155
18	113011	Jedeb	Nr. Amanuel	10.4	37.57	305	295
19	111003	Koga	@ Merawi	11.37	37.05	244	296
20	113036	L. Fettam	Nr. Galibed	10.48	37.02	757	758
21	111007	Megech	Nr. Azezo	12.48	37.45	462	508
22	112002	Mugher	Nr. Chanc	9.3	38.73	489	488
23	113026	Neshi	Nr. Shamb	9.75	37.25	322	327

24	112029	Robigumer	Nr. Lemi	9.75	39	887	897
25	112016	Shina	Nr. Adiet	11.25	37.5	111	111
26	113014	Temcha	Nr. Dembe	10.53	37.5	406	425
27	112040	Wenka	Nr. Istay	11.62	38.07	110	100

Table 9 compares two different areas in which the first (column 7) is from hydrology directorate (HD) and the second (column 8) is a delineated area.

3.4.2 Topography

The topography of the Abbay basin signifies two distinct features; the highlands, rugged mountainous areas in the center and eastern part of the basin and the lowlands in the western part of the basin. The river basin has the lowest elevation of 446 m and the highest elevation of 4246 m based on 30m DEM.

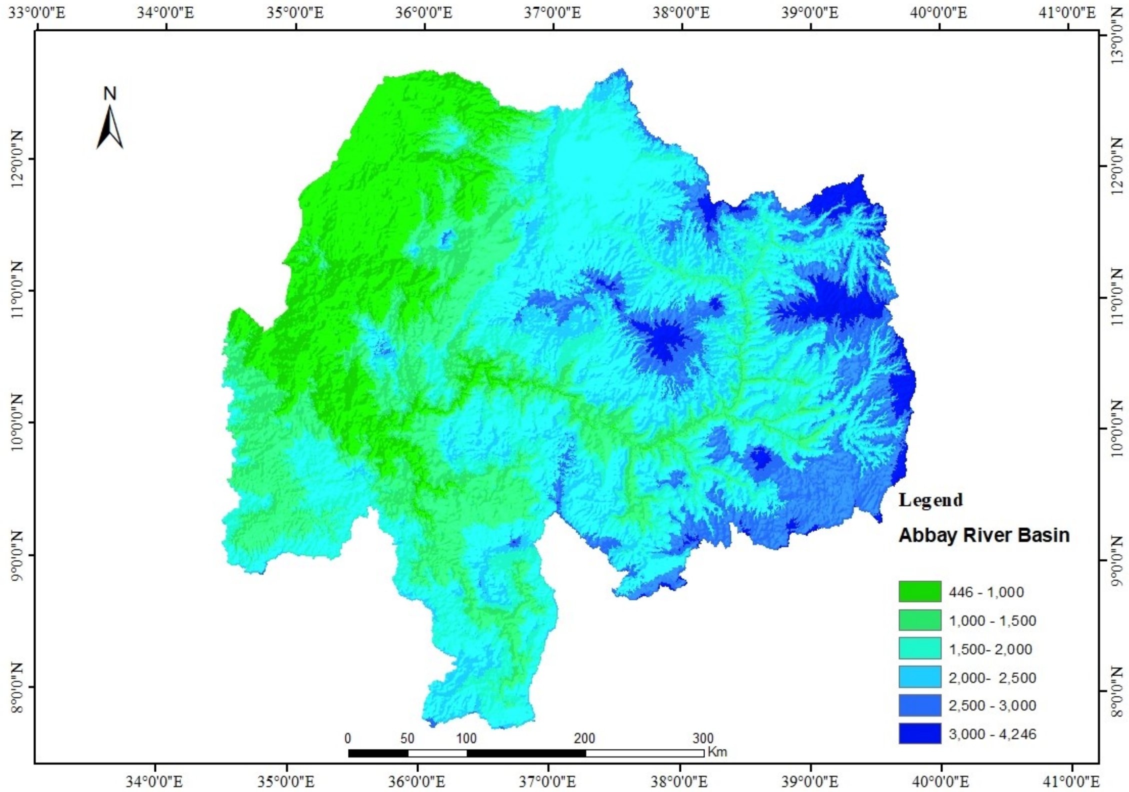


Figure 25 Topographical distribution of the Abbay River basin

The basin highlands extend from 1500 masl up to as high as 4264 masl, with a slope of greater than 20 percent in the eastern part. Whereas the basin lowlands flatten 1000 masl to 450 masl with a slope of less than 10 percent, in Dinder and Rahad sub-basins. The elevation and slope distribution of the Abbay river basin shows in Figure 25 and Figure 26 respectively.

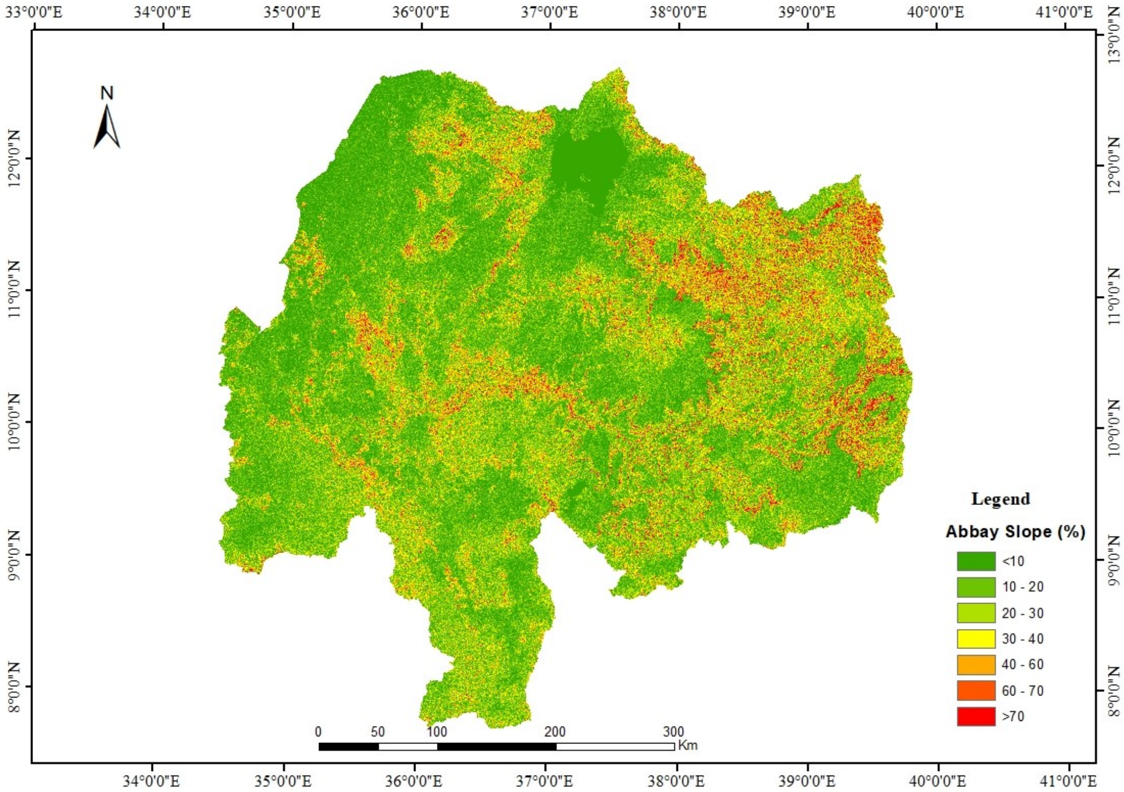


Figure 26 Land slope in percent distribution of the Abbay river basin

3.5 Explanatory variables

To perform regression analysis that estimates the response variable there must be explanatory variable/s that have influence in the formation of the response variable. There are different factors that have influences for runoff generation which are either climatic (rainfall, temperature, humidity, etc.) or physiographic variables (area, elevation, and slope, etc.). In this thesis, the explanatory variables are selected based on the easy availability of the variables and previous studies (Anisa Z. and Amanda E., 2018; Barbarossa et al., 2017; Burgers et al., 2013; Gebeyehu, 1989; Tara R. and Paulin C., 2013). Considering this reason one basic climatic variable (Mean Annual Rainfall) and three physiographic variables (Catchment Area, Average Slope, and Average Elevation) are selected. Generally, four Explanatory variables or predictors are used for analysis.

3.6 Regression analysis

Regression analysis is finding a relationship between the dependent and one or more independent variables. Sedighi, (2008) stated that "a model of the relationship is hypothesized, and estimates of the parameter values are used to develop an estimated regression equation" and if the model is passed a different test; the model can be used to forecast dependent variable for a value of different independent variables. Samprit C. and Ali S.H., (2012) stated that regression analysis regarded as iteration process that the output result will be repeatedly checked with diagnosis, criticize and validation process until the result is fittingly acceptable with the assumptions made primarily. This iterative nature of regression analysis showed in Figure 27 adapted from Samprit C. and Ali S.H., (2012).

The general standard output of regression analysis (Table 10) in statistical software including Microsoft Excel is the estimated regression coefficient, their standard error, t-test value in which that test the corresponding coefficient is zero, and p-value. Additionally, if the confidence level is selected the output will include upper and lower bounds of regression coefficients.

Regression analysis divided into two as simple and multiple linear regression analyses in which the main classification based on the number of predictor variables.

Table 10 Standard regression analysis output

Variable	Coefficients	Standard error	t-test	P value	Lower limit	Upper limit
Constant	β_o	-	-	-	-	-
X_1	β_1	-	-	-	-	-
X_2	β_2	-	-	-	-	-
. . .	. . .	. . .	. . .	. . .	. . .	. . .
X_n	β_n	-	-	-	-	-

Adapted from (Samprit C. and Ali S.H., 2012)

Simple linear regression

A simple linear regression model relates the response variable Y to a given single explanatory variable X through a straight-line equation,

$$Y = \beta_0 + X\beta_1 + \varepsilon$$

Where β_0 is called the constant-coefficient or the intercept, as it predicts the value of Y when X is equal to zero. β_1 is the slope that indicates the change of Y when X changes with one unit and ε are the random error which accounts for the differences between the regression model and the response sample points.

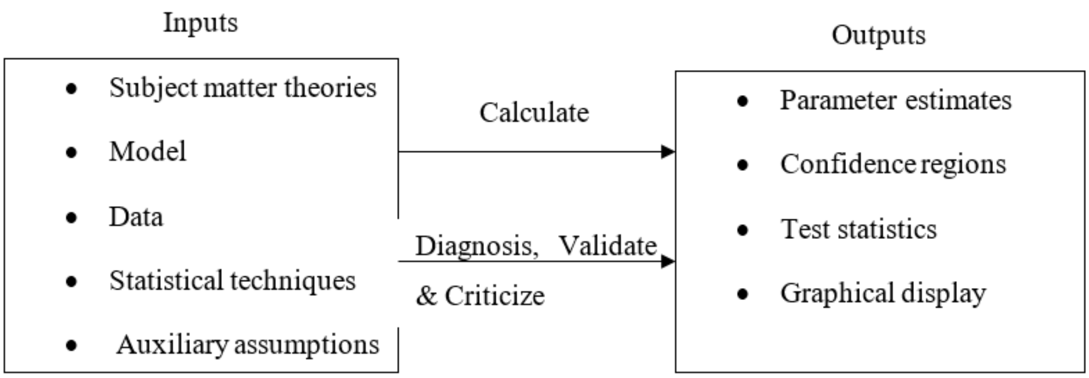


Figure 27 Iterative nature of regression analysis

Multiple linear regression

A regression model that involves more than one explanatory variable is called a multiple regression model. Which can be expressed as follows for responsive variable Y and to a given n number of explanatory variable X

$$Y = \beta_0 + X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + \cdots + X_n\beta_n + \varepsilon$$

Where β_0 is the constant-coefficient and similar to simple linear regression, it gives the value of Y when $X_1 = X_2 = ... = X_n = 0$. β is the regression coefficient, also called partial regression coefficient and it accounts for the influence of n numbers of explanatory variable X on the response variable Y after the coefficient and the variable have been adjusted to the other variables. ε is the random error.

Once the choice of predictor variables accomplished, the first steps in regression analysis are the determination of any linear relationships between selected explanatory variables (Gebeyehu, 1989). The reason behind this is that, in the case of multiple linear regression analysis, it will be difficult to know which variable affects the model if there is a relationship (high correlation coefficient) between predictors. As shown in Table 11 and Figure 28 there is no high correlation between predictors.

Table 11 Correlation matrix of variables

	Area	MAR	AE	AS	MAF
Area	1				
MAR	0.096	1			
AS	0.09	0.27	1		
AE	0.28	0.53	0.11	1	
MAF	0.96	0.015	0.10	0.20	1

Where: MAR is mean annual rainfall, AE is average elevation, AS is average slope and MAF is mean annual flow.

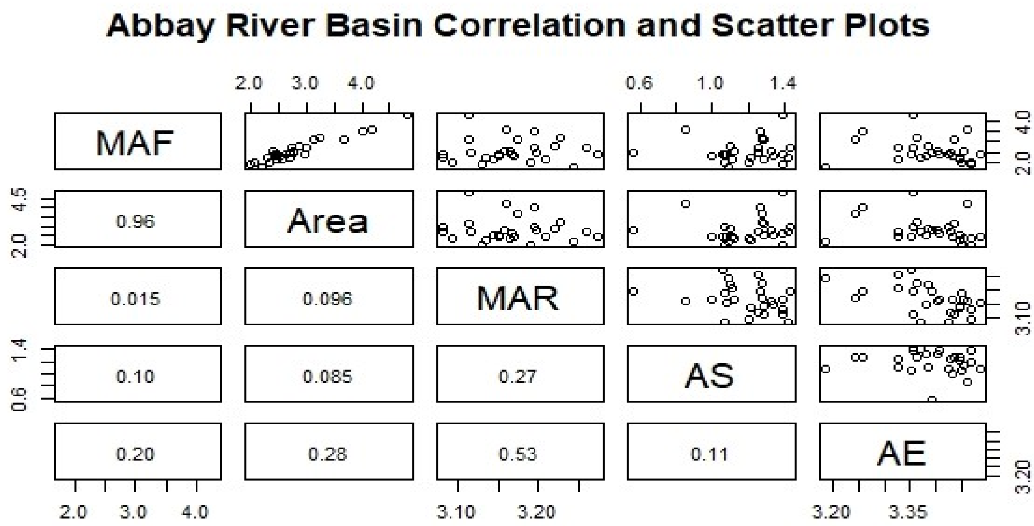


Figure 28 Abbay basin scatter plot with correlation coefficients of variables

3.7 Model development

In this thesis, two fundamental attempts were performed to establish empirical equations that will relate the dependent variable and independent variables. The first attempt was to get an empirical equation for each hydrometric station considering only rainfall as an independent variable. As stated below (Case -1) for this attempt two relationships (linear and nonlinear) between annual flow and annual rainfall were performed. Additionally, in Case-1 rainfall-runoff relationship estimated in terms of runoff coefficients. A second attempt (Case -2) was made in basin level with initially taking area (A) as the only explanatory variable and then adding mean annual rainfall. Finally, taking stepwise regression analysis whichever, gives good results the attempt was finished.

Case 1: relationship considering runoff as a function of rainfall only. Q = f {P}

In this case, linear and non-linear relationships between rainfall and runoff for each selected hydrometric station are established.

Linear relationships: in this form of relationship, any form of change in a given independent variable has the same form of change in the dependent variables. In other words, a straight-line relationship between dependent and independent variables. Mathematically, it can be expressed as the independent variable is multiplied by the slope coefficient, added by a constant, which determines the dependent variable.

In this relationship two forms of empirical equations, simple linear relationship and logarithmic transformation are tested.

i. Simple linear relationship $Q = \alpha + P * \beta$
ii. Transformed linear relationship $logQ = log\alpha + \beta * logP$

Where Q is annual flow (Mm^3), P is annual rainfall (mm), α is the constant coefficient or the intercept and β is the slope.

Non – linear relationships: change in any form of independent variables does not have any corresponding forms of constant change in a dependent variable. This relationship is the opposite of linear relationships. There are different types of non-linear relationships but for this study, only two of them power and polynomial second degree are tested.

i. Non-linear power form of relationship $Q = \alpha P^{\beta}$
ii. Polynomial second-degree non-linear relationship $Q = \alpha * P^2 + \beta * P + \gamma$

Where Q is annual flow (Mm^3), P is annual rainfall (mm), α, β, and γ are constant coefficients.

Furthermore, extra non-linear power function with a coefficient (c) in the form of $Q = \alpha(P - c)^{\beta}$ was established and if there is a change in model quality in the forms of R^2, R^2_{adj} and Ns then the equation was selected.

These stated forms of equations are tested using the Microsoft Excel solver add-in program. Solver finds an optimal (minimum) value for a formula in one cell called the objective function cell which is the sum of squared errors (SSE) and works with a group of cells, called decision variables in this case coefficients α, β, γ, and c. Solver adjusts the values in the decision variable (coefficients α, β, γ, and c) cells and produces the minimum SSE.

Runoff coefficients: The runoff coefficient (C) is a dimensionless number telling the ratio of the amount of runoff to the amount of rainfall. Runoff coefficient generally affected by land use, soil type, and slope of the catchment. It is a higher value for catchments relatively impervious, greater slope, and soil types such as clay and also lesser for catchments with vegetation to intercept runoff, lower slope, and soil that have higher infiltration rates such as soil contains sand. Estimating of runoff coefficient is important in different water resources development projects. For instance, for designing hydraulic structures, and for possible flood zone hazard delineation. A high runoff coefficient (C) value may indicate flash flooding areas during storms as water moves fast overland on its way to a river channel or a valley floor.

In this thesis runoff coefficient estimated by calculating the ratio of mean annual runoff (mm) to mean annual rainfall (mm) for each hydrometric station. This means that the percentage of rainfall that is not lost due to evapotranspiration, assuming storage negligible annually, and groundwater outflow out of the catchment as non-existent.

Case 2: relationship considering runoff as a function of combinations of explanatory variables whichever gives good results. MAF = f {A, MAR, AS & AE}

$$MAF = \alpha A^{\beta} * (MAR)^{\gamma} * (AS)^{\delta} * (AE)^{\mu}$$

Where: MAF is mean annual flow (Mm^3), MAR is mean annual rainfall (mm), AS is average slop (Percent), AE is average elevation (m), and α, β, γ, δ and μ are coefficients.

The above equation can be changed to linear by taking logarithm in both directions and it can be re-written as

$$logMAF = log\alpha + \beta * logA + \gamma * logMAR + \delta * logAS + \mu * logAE$$

The analysis starts with a single predictor (catchment area, A) i.e. simple linear regression and after that mean annual rainfall added. Lastly, any possible combination of independent variables gives better results. Hence, a stepwise regression analysis was done.

3.8 Model fitting

Model development task was done as stated in section 3.7Model development, though, it's not known which model is suitable, but before that coefficient needs to be determined. This task is done by selecting the most common ordinary least squares (OLS) as model fit; because, it results in an explicit equation, which facilitates interpretation and comparison with other studies.

3.9 Model criticism and selection

In regression analysis, the next step after parameter estimation is checking the adequacy of the model by performing the subsequent statistical tests.

Hypothesis tests

Hypothesis testing is a scientific process of testing whether or not the hypothesis is believable. The significance of the regression equation checked by testing the hypothesis of the null value of slope β_1 for simple regression and β_1 to β_n for multiple regression analyses. The first steps of a hypothesis test are development of null and alternative hypothesis in which followed by setting the significance level α (0.05 in this thesis) and calculating test statistics and probability value p-value. Finally, make decision-based on the calculated and tabulated tests and draw conclusion on it. So, in this study the hypothesis made to check the significance of the model is

H_0: $\beta_1 = 0$ for simple regression or $\beta_1 = \ldots\ldots = \beta_n = 0$ for Multiple regression

H_a: $\beta_1 \neq 0$ for simple regression or $\beta_k \neq 0$ at least one k; for Multiple regression

For simple linear regression, the test statistics used to check the significance of the regression equation are t-test. The test statistic t is constructed by comparing the slope of the estimated and the true regression line, as normalized by the standard error on the estimate of β_1 (Mauro Naghettini et al., 2017)

$$t = \frac{\hat{\beta}_1}{\sqrt{{S_e}^2 / \sum_{i=1}^{n}(X_i - \bar{X})^2}}$$

In this case, the calculated t value compared with critical t value which is read from the table using two-tailed (α/2) and n-2 degree of freedom. Consequently, the null hypothesis can be rejected if $|t| > t_{\frac{\alpha}{2}, n-2}$. Alternatively, the significance of the regression equation can be checked using an analysis of variance (ANOVA) as shown in Table 12.

Table 12 Analysis of Variance (ANOVA) table in regression analysis

Source of variation	Sum of squares	df	Mean square	F-Test
Regression	$SSR = \sum_{i=1}^{n}(\hat{y} - \bar{y})^2$	j	$MSR = \frac{SSR}{j}$	$F = \frac{MSR}{MSE}$
Residual	$SSE = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2$	n-j-1	$MSE = \frac{SSE}{n - j - 1}$	
Total	$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$	n-1		

Adapted from (Mauro Naghettini et al., 2017 and Samprit C. and Ali S.H., 2012)

Where j is a number of explanatory variables used in the regression analysis, df degree of freedom. Additionally, Microsoft excel adds one last extra column that gives the probability value of the F test. For simple linear regression, j is one, hence the degree of freedom will be n-2. Therefore, calculating a ratio of the mean squares due to regression (MSR) to residual mean squares (MSE) will give F -test.

$$F_o = \frac{MSR}{MSE} = \frac{SSR}{SSE/(n-2)}$$

If the calculated $|F_o| > F_{\alpha,1,n-2}$ the null hypothesis can be rejected.

For multiple linear regression significance of the regression equation test using the overall F-test. In this situation, the main difference with simple linear regression F- test is that due to the number of independent variables increment the degree of freedom of sum of squares due to regression and residual sum of squares are j and $n - j - 1$respectively; where j is the number of independent variables.

$$F = \frac{MSR}{MSE} = \frac{SSR/j}{SSE/(n-j-1)}$$

if $|F| > F_{\alpha,j,n-j-1}$the null hypothesis rejected. However, in this thesis for multiple linear regression analysis, the decision is not made by only F test, additionally, probability value (p-value) of each incorporated independent variable with a significance level (α) was compared and if p-value greater than α the hypothesis as well as the combination of independent variable is accepted.

A. Coefficient of determination (R^2) including adjusted(R^2_{adj})

One of the important tests for examining the quality of the model is with the coefficient of determination also called goodness of fit index or as the most commonly used R^2. This can be calculated by taking the ratio of the sum of squares due to regression and the total sum of squares (SST) or can be written as

$$R^2 = \frac{SSR}{SST} = \frac{\sum_i (Q_{Pri} - Q_{mn})^2}{\sum_i (Q_i - Q_{mn})^2}$$

Where Q_{pri} predicted mean annual flow (Mm^3)

Q_{mn} mean of observed data (Mm^3)

Q_i observed mean annual flow (Mm^3)

However, in multiple linear regression, the addition of the independent variable will increase the value of R^2-which can be leading to the wrong conclusion. To overcome this problem the modified or adjusted R^2 formula that only increase if the independent variable added improves the model.

$$R^2{}_{adj} = 1 - \frac{(1 - R^2)(N - 1)}{(N - j - 1)}$$

Where N number of observed data and j number of parameters

These coefficients range from 0-1 where 1 means a perfect fit between the observed values and the fitted ones, in other words, that the explanatory variable perfectly explains the response variable. However, for this study, acceptable values of R^2 and $R^2{}_{adj}$ are greater than 0.5.

B. Nash Sutcliffe (Ns)

The Nash-Sutcliffe efficiency (NSE) is a normalized statistic that determines the relative magnitude of the residual variance compared to the measured data variance (J.E.Nash and J.V.Sutcliffe, 1970). Nash-Sutcliffe efficiency indicates how well the plot of observed versus simulated data fits the 1:1 line. NSE = 1, corresponds to a perfect match of the model to the observed data. NSE = 0, indicates that the model predictions are as accurate as of the mean of the observed data, ∞ < NSE < 0, indicates that the observed mean is a better predictor than the model.

$$Ns = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_i (Q_i - Q_{pri})^2}{\sum_i (Q_i - Q_{mn})^2}$$

In this study model having Ns greater than 0.5 is accepted.

3.10 Model validation

Validation of the selected regression model in this thesis performed using 9 hydrometric stations (taking 30% of the total stations) considering as ungauged (pseudo-ungauged) catchments. These stations are mentioned in Table 13.

Table 13 Hydrometric station used for validation as pseudo-ungauged catchments

No	Station No	River	Site	Lat_DD	Lon_DD
1	112003	Abbay	@ Bahir Dar	11.6	37.4

2	114002	Angar	Nr. Nekemte	9.43	36.52
3	113029	Ardy	Nr. Metekel	10.95	36.52
4	112007	Beressa	Nr. D/Berhan	9.67	39.52
5	114001	Didessa	Nr. Arjo	8.68	36.42
6	111006	Gumara	Nr. B/Dar	11.83	37.63
7	113036	Lower Fettam	@ Galibed	10.48	37.02
8	111007	Megech	Nr. Azezo	12.48	37.45
9	112029	Robigumer	Nr. Lemi	9.75	39

CHAPTER 4 RESULT AND DISCUSSION

4.1 Results of Analysis

4.1.1 Rainfall-Runoff Relationship

4.1.1.1 An Empirical Equation for Each Hydrometric Station

One of the first tasks of this study is the formulation of the empirical equation for each selected station as specified in section 3.7. This analysis is done using Excel solver Add-in considering linear and non-linear relationships of independent variables. There are five different equation in which tested and their general forms are as follows

i. Non-linear power form of relationship $Q = \alpha P^{\beta}$
ii. Non-linear power form of relationship with coefficient $Q = \alpha(P - c)^{\beta}$
iii. Simple linear relationship $Q = \alpha * P + \beta$
iv. Transformed linear relationship $logQ = log\alpha + \beta * logP$
v. Polynomial second-degree non-linear relationship $Q = \alpha * P^2 + \beta * P + \gamma$

Hence, Excel solver Add-in is used for the analysis, initial trial values were given and the first trial values of coefficients got but the process proceeds until optimal values of coefficients are reached. As a threshold value of 0.5 is stated in section 3.9 to select the model for developed forms of equations using R^2, $R^2_{adj,}$ and Ns.

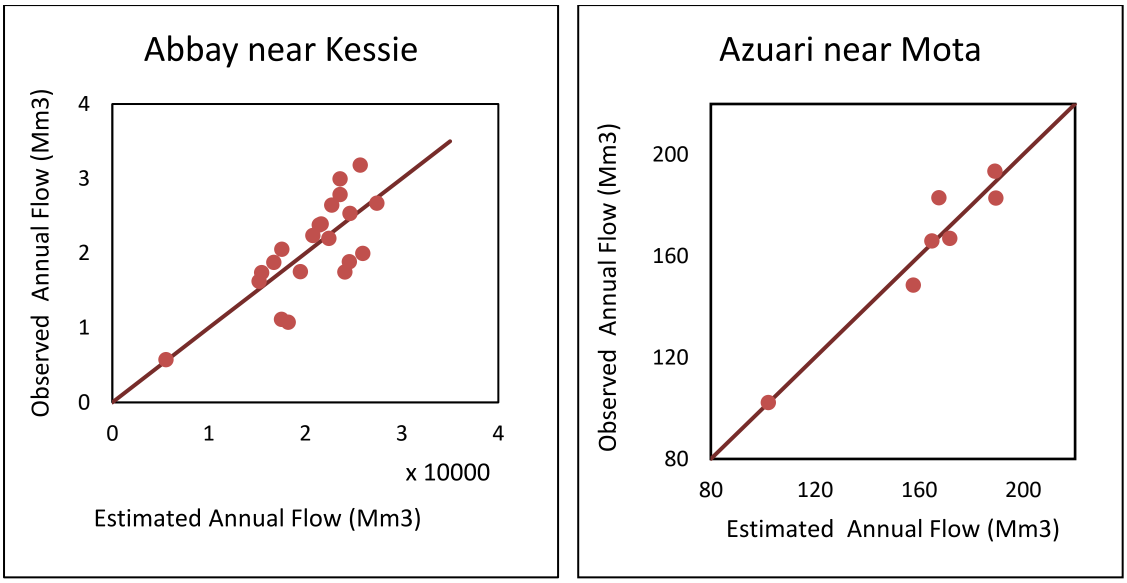
Generally, only 15 hydrometric stations pass the requirements. Table 14 summarizes successful stations with their best-fit forms of relationship including the estimated coefficient, corresponding coefficients of determination (R^2), and Nash Sutcliffe (Ns). Note that more than half of the forms of relationship are Power with a coefficient greater than 500mm except Jedeb near Amanuel gauge (381.77mm) and with three log-transformed, and four power rule forms. Any forms of relationships stated above were not show acceptable results for the other 12 hydrometric stations. The full tabular results of all hydrometric stations can be referring in Appendix –B.

Table 14 Successful hydrometric stations in rainfall-runoff relationships annually

Station	Forms	α	β	γ	R2	Ns
Abbay Kessi	Transformed	1.4E-05	2.94		0.69	0.69

Andassa	Transformed	7.60E-05	2.06		0.5	0.5
Anger	Power W cof	3.6E-07	3.27	622.99	0.66	0.64
Azuari	Power W cof	71.93	0.17	1238.06	0.93	0.93
Bello	Power	0.0012	1.64		0.51	0.51
Birr	Power W cof	6.9E-06	2.61	517.88	0.77	0.77
Dura	Power W cof	1.24	0.86	588.06	0.54	0.54
Gilgel Abbay	Power W cof	762.78	0.14	1246.14	0.51	0.51
Jedeb	Power W cof	1.90E-06	2.68	381.77	0.72	0.69
Megech	Power W cof	10.8	0.49	916.21	0.57	0.57
Mugher	Power	0.049	1.204		0.5	0.5
Neshi	Power	1.2E-07	2.928		0.69	0.7
RobiGumero	Transformed	0.00177	1.66		0.63	0.63
Shina	Power W cof	4.13E-08	3.120	564.85	0.73	0.58
Temecha	Power	5.6E-05	2.1		0.62	0.62

Figure 29 shows the result of the estimated annual flow plotted against the observed annual flow for stations that have a higher coefficient of determination (R^2) and Nash Sutcliffe (Ns) for sample named stations in Table 14.



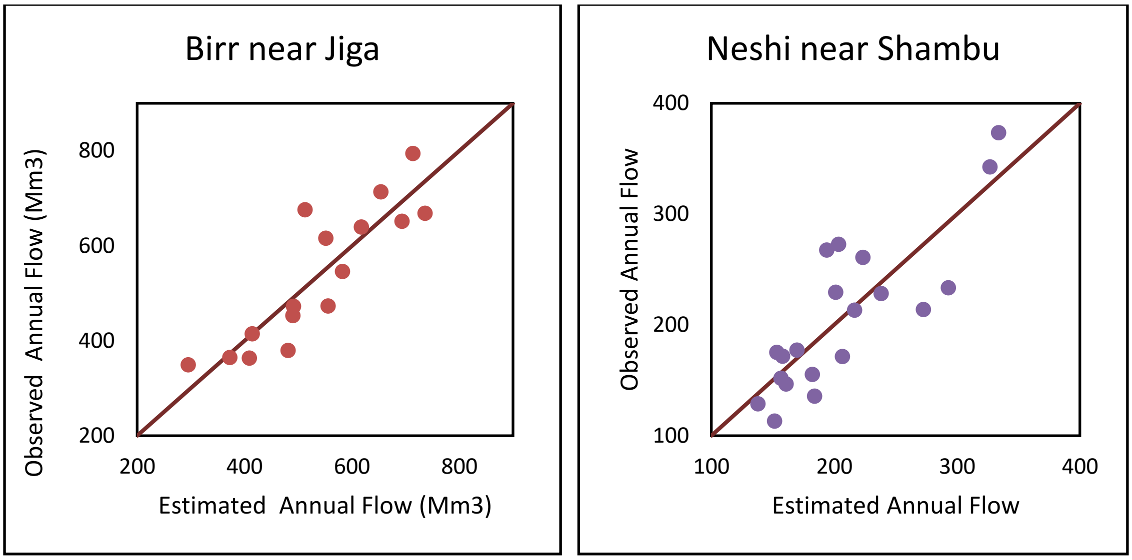


Figure 29 Comparison of Estimated and Observed annual flow for sample selected stations

4.1.1.2 Runoff coefficient

Runoff coefficient estimated by simply calculating the ratio of mean annual runoff (mm) with mean annual rainfall (mm) for each hydrometric station. The summarized results of runoff coefficients are shown in Table 15 below.

Table 15 Runoff coefficient

Station Name	C	Station Name	C
Abbay Nr. Kessi	0.23	Gudla	0.98
Abbay @ BDR	0.19	Gumara	0.6
Andassa	0.31	Hoha	0.20
Angar	0.26	Jedeb	0.61
Ardy	0.59	Koga	0.30
Azuari	0.58	Lower Fettam	0.51
Bello	0.45	Megech	0.31
Beressa	0.38	Mugher	0.40
Birr	0.36	Neshi	0.44
Chemoga	0.32	Robigumer	0.23
Didessa	0.20	Shina	0.48
Dura	0.54	Temcha	0.42
Gelegel Abbay	0.61	Wenka	0.56
Guder	0.44		

4.1.2 Regression Analysis in basin level

As stated in section 3.7 the second case in this study is a model developed using regression analysis. Here, the analysis starts with simple regression taking catchment area (A) as a predictor variable and proceed by adding mean annual Rainfall i.e. multiple linear regression. Lastly, stepwise linear regression analysis is done. The diagnosis of regression analysis is also performed.

4.1.2.1 Regression analysis Area (A) as the only Predictor

This analysis area is used as the only predictor variable that estimates the response variable mean annual flow (Q).

The logarithmic transformation forms of the simple linear regression equation shown below are used.

$$logMAF = log\alpha + \beta * logA$$

Table 16 Regression analysis result Area as the only predictor

Multiple R^2	R^2	Nash Sutcliffe	Standard error
0.96	0.93	0.93	0.1577

Based on Table 16 one can easily decide that the relationship between the dependent and independent variables is excellent.

Table 17 Coefficient table for Area as only predictor variable

Coefficient table						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	0.0957	0.1699	0.5635	0.5809	-0.2644	0.4559
Area	0.8930	0.0621	14.3697	1.5E-10	0.7612	1.0247

Table 17 and Table 18 summarizes the coefficients, hypothesis test, and statistical tests of mean annual rainfall and catchment area regression analysis. As it is seen t-test of slope β is 14.37 which is greater than $t_{(0.025,17)}$ which is read as 2.11 from the student's t-table and the corresponding p-value is less than the adapted significance level 0.05. Additionally, F-test in the ANOVA table shows it is much greater than $F_{(0.05,1,17)}$ which is read as 4.45 from the F-distribution table and the corresponding p-value is less than the adapted significance level 0.05.

Additionally, a comparison was made on estimated and observed values graphically as shown in Figure 30.

Table 18 Analysis of variance (ANOVA) table for Area as the only predictor variable

ANOVA Table					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	5.1378	5.1378	206.49	1.45E-10
Residual	16	0.3981	0.0249		
Total	17	5.5359			

4.1.2.2 Regression analysis Area and Mean Annual Rainfall as a predictor

Multiple linear regression analysis starts by adding mean annual rainfall to the starting regression analysis of the simple linear regression area as the only predictor variable. The log-transformed form of this regression analysis is

$$MAF = log\alpha + \beta \log(A) + \gamma \log(MAR)$$

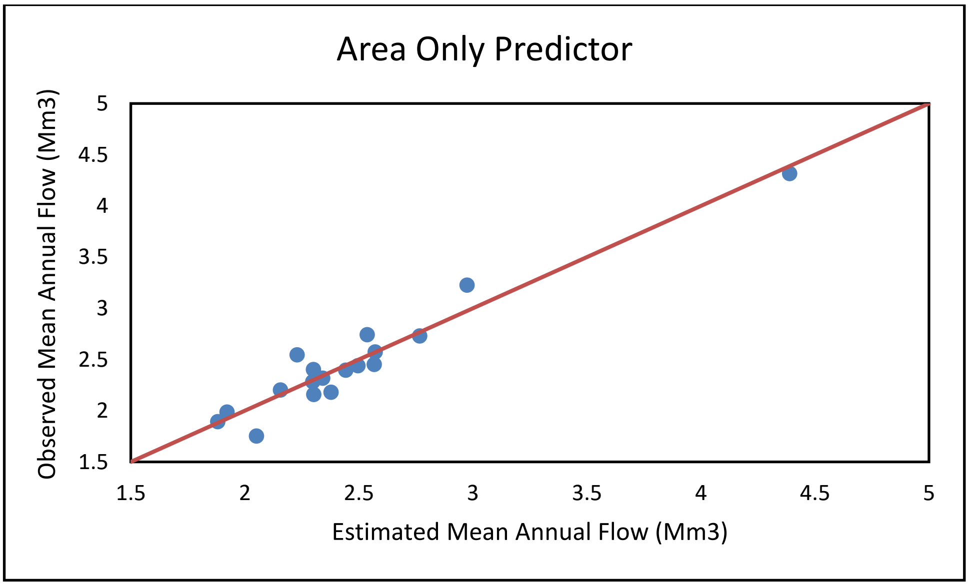


Figure 30 Comparison of Estimated versus Observed MAF using Area only as a predictor

Table 19 Regression analysis results using Area and Mean Annual Rainfall as predictors

Multiple R^2	R^2	R^2_{adj}	Nash Sutcliffe	Standard error
0.96	0.93	0.92	0.93	0.1623

Comparing Table 16 and Table 19 as expected slightly increment in four-digit of coefficients of determination (R^2). In addition, for the second case adjusted coefficients of determination (R^2_{adj}) are considered. There was also increasing in standard error detected.

Table 20 Coefficient table for Area and Mean Annual Rainfall as predictor variables

Coefficient table						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-0.9638	3.0620	-0.3148	0.7573	-7.4902	5.5626
Area	0.8968	0.0649	13.8217	6.1E-10	0.7585	1.0351
MAR	0.3307	0.9541	**0.3466**	**0.7337**	-1.7030	2.3644

Even though, from ANOVA table (Table 21) F-test shows null hypothesis test is rejected coefficient table (Table 20) exhibit that the added explanatory variable MAR is not significant because corresponding t-test value and p-values are less than $t_{(0.025,16)} = 2.12$ and significance level 0.05 respectively.

Table 21 ANOVA table for Area and Mean Annual Rainfall as the predictor variables

ANOVA Table					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	5.1409	2.5705	97.6263	2.51E-09
Residual	15	0.3950	0.0264		
Total	17	5.5359			

4.1.2.3 Regression analysis with the most suitable predictor variables

The last regression analysis was finding the most appropriate predictor variables for the basin. Back-ward elimination stepwise regression analysis was used to find the best fit model. Therefore, all independent variables were initially used and if there is a violation of the hypothesis test in p-value or t-test the corresponding independent variable that violates will eliminate from the analysis and using the remaining independent variables the analysis will be repeated until the best fit that satisfies achieved.

Table 22 Regression analysis result using all independent variables

Multiple R^2	R^2	R^2_{adj}	Nash Sutcliffe	Standard error
0.98	0.97	0.96	0.97	0.1130

Unfortunately, in this thesis regression analysis using all predictor variables initially satisfy t-test, F-test, p-value (except average slope) and have very high values of R^2, $R^2_{adj,}$ and Nash Sutcliff and also the low value of standard error as shown in Table 22, Table 23, and Table 24.

Table 23 Coefficient table of all independent variables

Coefficient table						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-19.6736	4.9082	-4.0083	0.0015	-30.2771	-9.0701
Area	0.9713	0.0504	19.2564	6.1E-11	0.8623	1.0803
MAR	3.5216	1.0054	3.5027	0.0039	1.3496	5.6936
AS	0.2144	0.1440	**1.4888**	**0.1604**	-0.0967	0.5255
AE	2.3947	0.5754	4.1615	0.0011	1.1515	3.6378

Table 24 ANOVA table for all independent variables as the predictor

ANOVA Table					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	5.3699	1.3425	105.12	9.20E-10
Residual	13	0.1660	0.0128		
Total	17	5.5359			

Next, eliminating the average slope and again regression analysis with the remaining independent variable was done. Table 25 shows a minor dropping of R^2 and Nash Sutcliff and increment of the standard error of the model. t-test, F-test, p-value are satisfied as shown in Table 26 and Table 27.

Table 25 Regression analysis result using Area, MAR, and AE as predictors

Multiple R^2	R^2	R^2_{adj}	Nash Sutcliffe	Standard error
0.98	0.96	0.96	0.96	0.1178

Table 26 Coefficient table of Area, MAR, and AE as predictors

Coefficient table						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-17.5316	4.8922	-3.5836	0.0030	-28.024	-7.0388
Area	0.9815	0.0521	18.8359	2.4E-11	0.8697	1.0932
MAR	3.0777	1.0010	3.0746	0.0082	0.9308	5.2246
AE	2.2463	0.5909	3.8018	0.0019	0.9791	3.5136

Table 27 ANOVA table for Area and Mean Annual Rainfall as the predictor variables

ANOVA Table					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	5.3415	1.7805	128.28	2.03E-10
Residual	14	0.1943	0.0139		
Total	17	5.5359			

This combination of catchment characteristics is statistically significant (t-test > t-cr and p-value < 0.05). Observed and estimated mean annual flow using area, mean annual rainfall, and the average elevation is shown in Figure 31.

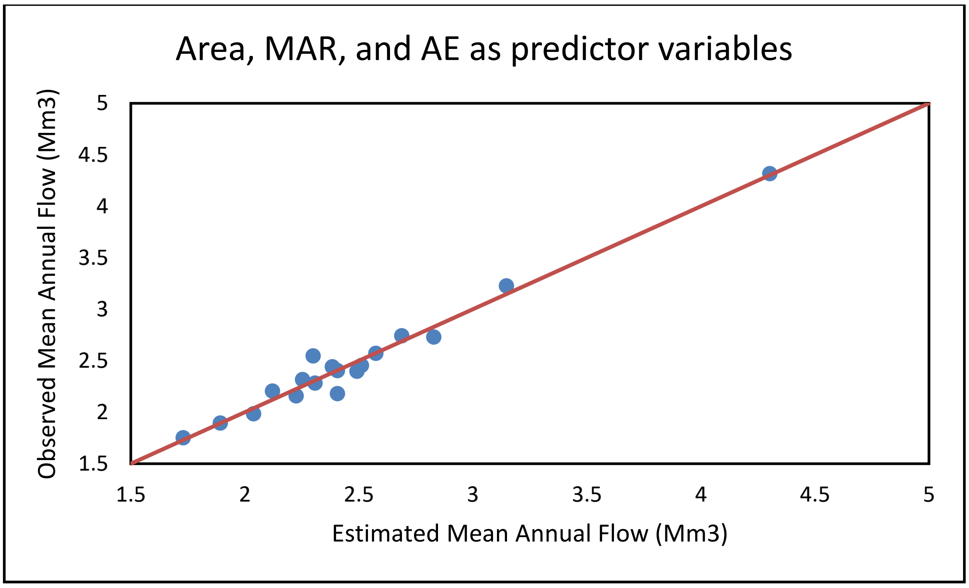


Figure 31 Estimated versus observed MAF of Area, MAR, and AE as a predictor

4.1.3 Validation results

The validation analysis made only on two equations that pass statistical tests; these are the simple linear regression that Area as the only predictor and multiple regression Area, Mean Annual Rainfall, and Average Elevation as predictors are the equations. The summarized results of regression analysis are shown in Table 28. The comparison was made on how much the estimated and observed values on each station are related in terms of magnitude and Graphically (Figure 32). Using the coefficients in Table 28 the observed and estimated mean annual flow for each equation was estimated. The first equation has well estimated in the Abbay @ BahirDar and Gumara stations compared with the second equation. However, for the remaining seven stations the second equation estimates excellently well with the first equation.

Table 28 Summarized coefficients of regression analysis

Predictor	α	β	γ	δ	η
Area	$10^{0.096}$	0.89			
A, MAR & AE	$10^{-17.53}$	0.98	3.08		2.25

The coefficient of determination and Nash Sutcliffe were calculated for each equation and the first equation gives more acceptable results than the second one. This is due to that the estimation of the second equation for Abbay @ BahirDar is almost three times the observed one and the first equation tells that two catchments that have the same area in the basin have the same mean annual rainfall which doesn't give sense, at least unless both of the catchments are in the same rainfall distribution parts of the basin. Therefore, considering these reasons and seeing Figure 32 the second equation (A, MAR, & AE as predictors) selected as a validated equation. The validation statistics are shown in Table 29.

Table 29 R^2 and Ns of Validation results

Predictor	Area	A, MAR, & AE With Abbay @ BDR	A, MAR, & AE Without Abbay @ BDR
R^2	0.97	0.80	0.96
Ns	0.80	0.49	0.96

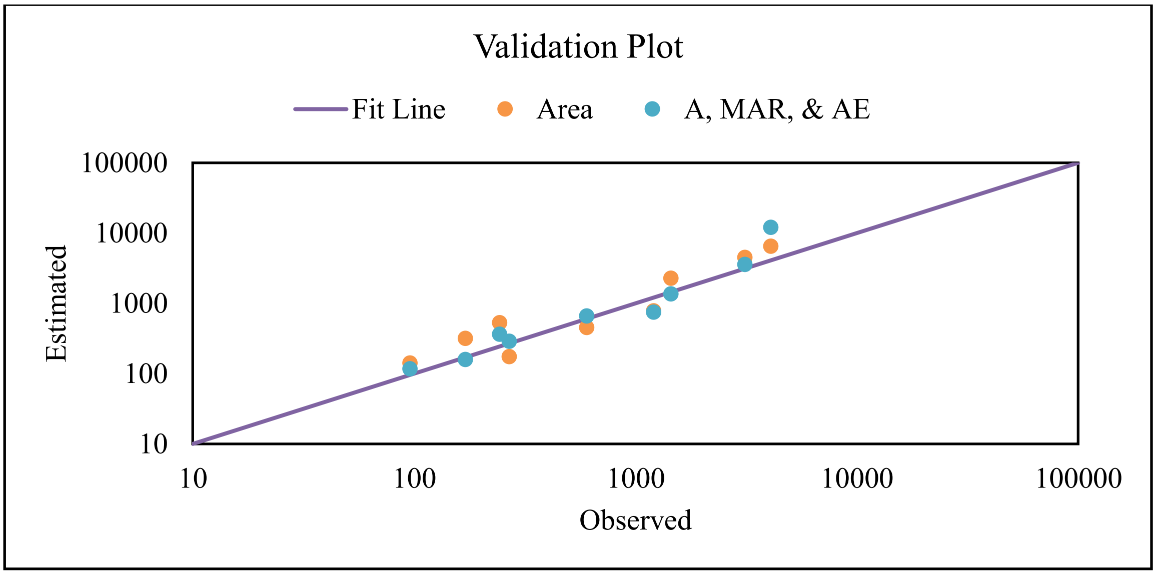


Figure 32 Validation plots of the Observed and Estimated Mean Annual Flow

4.2 Discussion

4.2.1 Rainfall-Runoff relationship

Water is one of the essential needs of human life in such a way that used for day to day usage as well as future benefits by storing. Nevertheless, the benefit that extracted from the water was not possible due to a limited number of measuring gauges in the watersheds. Even though they are recorded they are not consistent. This thesis has resulted in two vital attempts in the Abbay river basin. The first one is to identify and estimate the rainfall-runoff relationships for the selected forms of equations and runoff coefficients at each hydrometric station respectively. The second attempt is the development of a regression model that can be used to estimate mean annual flow (MAF) at ungauged catchments in the Abbay river basin.

In the development of an empirical equation for rainfall-runoff relationships considering linear and non-linear forms of relationships of annual flow with annual rainfall the successful that satisfy statistical requirements, hydrometric stations are 15 in number. All of the selected forms of relationships are non-linear and this indicated that the theoretical explanation of rainfall and runoff relationships is not linear is valid. The remaining 12 hydrometric stations do not satisfy the statistical requirements, it is possible to speculate that this might be due to that the selected forms of relationships are not enough to explain. Runoff coefficients that show in Table 15 also relate mean annual rainfall and mean annual flow at each gauge stations. The numerical values tell the effect of land use, soil, slope, groundwater level, and rainfall intensity in the considered catchment. Seleshi B. Awulachew et al., (2011) summarize the statistics of major sub-basin of the Blue Nile basin, runoff coefficient was one of the statistics and one station Abbay at Bahir Dar totally the same (0.19) with the sub-basin lake Tana. Additionally, Abbay nr Kessie and Dedissa Nr Arjo stations they are almost the same, However, eight hydrometric stations show more than 50% of the change of rainfall amount into a runoff. From theme, Gilgel Abbay and Gumara hydrometric stations are the known catchments as studied by M. Dessie et al., (2014) 65% of the flow in Gumara catchment appears in the form of interflow and 40-48% of flows are baseflow for Gilgel Abbay catchment. Furthermore, they also find out that the amount of direct runoff from the total runoff in these two catchments is very low 18-19% for Gumara and 20% for Gilgel Abbay. Moreover, SMEC, (2008) reported that there is substantial underground inflow into the catchments of the Gilgel Abbay and the Gumara through the thick basalts from elevated areas around the old volcanoes Guna and Choke. Gudla near Dembecha hydrometric station shows a runoff coefficient of almost unity (0.98) which is unrealistic and very far from the sub-basin South Gojam (0.18).

4.2.2 Mean Annual Flow Analysis

This analysis found evidence for the relationship between flow with catchment characteristics in the Abbay river basin. The result section afterward section 4.1.2 shows the ordinary least square regression model of mean annual flow (MAF) with four independent variables (Area, Mean Annual Rainfall, Average Slope, and Average Elevation). This analysis starts with Area as the only predictor and gives 0.93 in both coefficients of determination (R^2) and Nash Sutcliffe (Ns) with a standard error of 0.1577. The significance of the model also shows acceptable with a coefficient table (Table 17) and ANOVA table (Table 18), moreover, observed versus estimated mean annual flow shows graphically how the model is fit. The second attempt is additional of mean annual rainfall as an independent variable and which does not show any significant improvement of the model rather than the increment of standard error (0.1623), in addition, to failing to satisfy the hypothesis test. As we see in Table 20 the exponent of the Area does not change but the sign of the intercept is opposite with area as the only predictor and the exponent of mean annual rainfall is less than 0.5.

Lastly, the most suitable model is selected using a stepwise regression model with backward elimination by starting from taking of all independent variables as predictors. In this time R^2, R^2adj, and Ns increase 0.97,0.97, and 0.96 and standard error decrease (0.1130). However, in Table 23 Average slope does not satisfy the hypothesis test ($p > \alpha$); and it shows the most important independent variable is Area, Average Elevation, and Mean Annual Rainfall respectively based on standard error. Next, the analysis proceeds by eliminating the Average Slope and it shows that the Area, Mean Annual Rainfall, and Average Elevation are statistically significant with 0.96 in R^2, R^2adj, and Ns. In this regression equation, the exponent of predictors shows their relationships with the mean annual flow. For the area the exponent is almost one (0.98), hence, the relationship between the size of the catchment and mean annual flow is linear or in other words unit increase in the area contributes a unit volume of water to the river channel. Mean annual rainfall has exponents greater than one (3.08) which reflects non-linear relationships between them and strengthens the first objective of this thesis. Also, the exponent of average elevation is also greater than one that tells there are no linear relationships between them.

The comparison of previous study results on the basis of the coefficient of determination and the number of independent variables incorporated. Vogel et al., (1999) for Area as the only predictor R^2 obtained 0.91 and with all the variables as predictors R^2 increased to 0.99. Additionally, the number of geomorphological and climate variables was more associated with this study and they used a maximum of five predictors to develop the most suitable mean annual flow that in almost all regions catchment area was inclusive. D. M. Papamichail et al., (2002) select the best-fitted model that include Area, Annual rain, and length of a stream with R^2 of 0.78. Hence, their result is not much greater than this study. The other similar study report was the master plan project of Rift Valley Lakes basin HALCROW G. L., (2007) for Area as the only predictor they obtained R^2 of 0.82 and used natural logarithms transformation for Area, Annual Average Rainfall, and Altitude. For transformed independent variables Area as the only predictor, R^2 was 0.69 which is less than this study. The best-fitted appropriate model selected to estimate mean annual flow for ungauged catchments in Rift Valley Lakes River basin was R^2 of 0.72 and R^2_{adj} of 0.65 that includes Area and Annual Average Rainfall. Whereas in this study the regression model that includes Area and Mean Annual Rainfall is R^2 of 0.93 and R^2_{adj} of 0.92 and this model is not selected as best or fit cause it does not pass hypothesis test.

Moreover, Burgers et al., (2013) founded R^2 of 0.40 for Area as the only predictors and they obtained an increment of R^2 of 0.56, while, in this study, the inclusive of Mean Annual Rainfall do not improve the model rather than the increment of standard error. The exponent of the Area looks similar to this study of 0.86. The above-mentioned studies were almost used similar climate as well as physiographic variables with this study, however, Tran et al., (2015) include geomorphologic and anthropogenic variables for 533 catchments. They obtained a better result than this study. Barbarossa et al., (2017) developed a global regression model using catchment area, mean annual rainfall, mean annual temperature, mean elevation and mean slope as the predictor for 1885 catchments. They obtained R^2 of 0.89 which very less compared with this study.

As mention in the above catchment area, the mean annual rainfall, and the average elevations are statistically significant next to the area as the only predictor. However, the validation analysis shows that multiple regression equation that includes the three catchment characteristics is the best equation capable of estimating the mean annual flow in ungauged catchments in the basin. Therefore, the equation looks like

$$MAF = 10^{-17.53} A^{0.98} * MAR^{3.08} AE^{2.25}$$

Where: A is catchment area (Km^2), MAR is mean annual rainfall (mm), and AE is average elevation in (m).

The limitations of the present studies naturally include exclusive of the independent variables which might have an influence on the formation of runoff through the catchment. These are the soil characteristics, land use/land cover, shape of the catchment and temperature, etc. It suffers from the same limitations as other local studies associated with the quality and number of hydrometeorological data.

CHAPTER 5 CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

This thesis aimed to identify the rainfall-runoff relationship and the development of empirical equations that can be used for ungauged catchments in the Abbay river basin. The relationship of rainfall-runoff examined in terms of developing linear and non-linear forms of an equation between annual rainfall and annual flow. Additionally, the relationships were tested by estimating runoff coefficients. Climate and physiographic characteristics were used to develop the empirical equation that satisfies statistical tests.

Consequently, based on the methodology followed and the results attained, the following conclusions are pinched:

- Through the development of relationships between annual flow versus annual rainfall, among the total 27 hydrometric stations, 15 catchments were passed the minimum requirement of R^2 and Ns with non-linear forms of equations and the remaining 12 hydrometric stations were failed. The failure of these stations can be due to the selected forms of the equation are not enough to explain the relationships of rainfall and runoff. Additionally, relationships of rainfall-runoff described by estimating the runoff coefficient at each selected hydrometric station and the result obtained were greater than 0.5 for 8 stations and basically two stations namely Gumara and Gilgel Abbay the amount of runoff was not only from the rain alone the interflow and base flow contribute much respectively.

- The developed regression model that capable of estimate the mean annual flow for ungauged catchments includes catchment area (Km^2), mean annual rainfall (mm), and average elevation (m) as predictors that statistically significant with values of 0.96 for all R^2, R^2_{adj}, and Ns.

- The validation using nine hydrometric stations shows that the selected developed model is capable of estimate the mean annual flow for ungauged catchments in the Abbay river basin with respect to R^2 and Ns of 0.96 and using validation plot (Figure 32).

Therefore, it is possible to use the developed empirical equation in the water resources development project for ungauged catchments in the Abbay river basin.

5.2 Recommendations

To better understand the implications of these results, future studies could address the following listed ideas.

- The number of meteorological stations within and outside the basin should be more than used in this study to improve the model quality.
- Selected hydrometric stations do not cover all the basin areas, as it known northwest side of the basin are Rahad and Dinder sub-basins that join the Abbay river outside the Country. Hence, a recommendation is given to add hydrometric stations specifically, in Beles, Wenbera, Beshilo, and Welaka sub-basins.
- It is also recommended to include geomorphological and anthropogenic catchment characteristics.

REFERENCES

Abebe S.G., S. A. (2013). Basin Regionalization for the Purpose of Water Resource Development in a Limited Data Situation: Case of Blue Nile River Basin, Ethiopia. *ResearchGate*.

Anisa Z. and Amanda E. (2018). *Regional Regression Models of Mean Annual Streamflow and Design Flow in a Tropical Region of Colombia.*

Barbarossa, V. H. (2017). Developing and testing a global-scale regression model to quantify mean annual streamflow. *Journal of Hydrology*, 479–487.

BCEOM, F. E. (1999). *Abbay River Basin Integrated Development Master Plan Project, Section II - Sectorial Studies, Volume III- Water Resources.* Addis Ababa, Ethiopia: Ministry of Water Resources.

Belete B, Y. S. (2014). Chapter 6 Surface Water and Groundwater Resources of Ethiopia: Potentials and Challenges of Water Resources Development. In W. A. Assefa M. Melesse, *Nile River Basin, Ecohydrological Challenges, Climate Change and Hydropolitics* (pp. 97-117). New York: Springer International Publishing.

Berhane, M. (2013). *Estimation of Monthly Flow for Ungauged Catchment (Case Study Baro - Akobo basin), MSc Thesis.* Addis Ababa , Ethiopia: Addis Ababa University.

Bhagat NK. (2017). Rainfall-runoff co-relationship for lower mahi basin, India. *International Journal of Hydrology*, 127–130.

Burgers, H. S. (2013). Size relationships of water discharge in rivers: scaling of discharge with catchment area, main-stem length and precipitation. *Hydrological Process* , 5769–5775.

Chekole Tamalew and Abdella Kemal. (2016). Estimation Of Discharge For Ungauged Catchments Using Rainfall-Runoff Model In Didessa Sub-Basin: The Case Of Blue Nile River Basin, Ethiopia. *International Journal Of Innovations In Engineering Research And Technology*, 62-72.

D. M. Papamichail, P. E. (2002). Multiple Regression Models for Predicting Annual Runoff in Ungauged Catchments. *ResearchGate*.

Devi, G. K. (2015). A Review on Hydrological Models. *Aquatic Procedia, 4*, 1001 – 1007.

Douglas C. M., E. A. (2012). *Introduction to linear regression analysis.* Hoboken, New Jersey.: John Wiley & Sons, Inc.,.

Gebeyehu, A. (1989). *Regional Flood Frequency Analysis, Dissertation.* Stockholm.

Gebeyehu, H. (2013). *Remote Sensing and Regionalization for integrated water resources modeling, in Upper and middel Awash River basin.* Enschede,Netherlands: Msc Thesis.

HALCROW. (1989). *Master Plan for the Development of Surface Water Resources in the Awash Basin, Final Report, Volume - 4.* Addis Ababa, Ethiopia: Ethiopian Valleys Development Studies Authority.

HALCROW, G. L. (2007). *Rift Valley Lakes Basin Integrated Resources Development Master Plan Study Project, Draft Phase 1 Report Part II, Volume -5.* Addis Ababa, Ethiopia: Ministry of Water Resources.

J. Odry and P. Arnaud. (2017). Comparison of Flood Frequency Analysis Methods for Ungauged Catchments in France. *Geosciences*, 1-24.

J.E.Nash and J.V.Sutcliffe. (1970). River Flow Forcasting Through Conceptual Models Part I - A Discussion of Principles. *Journal of Hydrolgy*, 282-290.

Jan Sitterson, C. K. (2017). *An Overview of Rainfall-Runoff Model Types.* Athens, Georgia: EPA,United States Enviromental Protection Agency .

LAHMEYER, I. a.-B. (2007). *Genale Dawa River Basin Integrated Resources Development Master Paln Study, Final Report, Part II, Volume - II.* Addis Ababa, Ethiopia: Ministry of Water Resources.

M. Dessie, N. E. (2014). Analyzing runoff processes through conceptual hydrological modeling in the Upper Blue Nile Basin, Ethiopia. *Hydrology and Earth System Sciences*, 5149-5167.

M. Goswami, K. O. (2007). Development of regionalisation procedures using a multi-model approach for flow simulation in an ungauged catchment. *Journal of Hydrology*, 517-531.

M. Sivapalan, K. T. (2003). IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an Exciting Future for the Hydrological Sciences. *Hydrological Sciences Journal*.

Mauro Naghettini, A. T. (2017). *Fundamentals of Statistical Hydrology.* Switzerland: Springer International Publishing.

MoWIE, R. a. (2018). *Water, Irrigation and Electricity Researchable Issues and List of Research Outputs/Index.* Addis Ababa: Publication No. 5th.

NOHRSC, N. O. (2019, 10 28). *National Weather Service*. Retrieved from https://www.nohrsc.noaa.gov/technology/gis/uhg_manual.html

P.K. Singh, S. M. (2014). A review of the synthetic unit hydrograph: from the empirical UH to advanced geomorphological methods. *Hydrological Sciences*, 239-261.

Pasquale C., G. C. (2006). Regional Models for the Estimation of Streamflow Series in Ungauged Basins. *ResearchGate*, 789–800.

Ralf M., G. B. (2006). Regionalization methods in rainfall–runoff modelling using large catchment samples. *IAHS Publication*, 117-125.

Rao .A.R. and Srinivas V.V. (2008). *Regionalization of Watersheds, An Approach Based on Cluster Analysis.*

S.K. Jain and V.P. Sing. (2003). *Water Resources Systems Planning Management. .* Amsterdam:: Elsevier Science B.V.

Samprit C. and Ali S.H. (2012). *Regression Analysis by Example Fifth Edition.* Hoboken, New Jersey: John Wiley & Sons.inc.

Sedighi, A. (2008). *A Linear Regression Model For Predicting Stream Response Time In Karst Watersheds Using Dems, Dissertation. .*

Seleshi B. Awulachew, F. D. (2011). Chapter 14 Hydrological Water Availability, Trends and Allocation in the Blue Nile Basin. In A. M. Melesse, *Nile River Basin, Hydrology, Climate and Water Use* (pp. 283 -296). New York: Springer International Publishing.

SMEC. (2008). *Hydrological Study of the Tana -Beles Sub-basins, Surface water Investigations , Project Number: 5089018.* Addis Ababa, Ethiopia: MoWIE.

Tara R. and Paulin C. (2013). Streamflow Prediction in Ungauged Basins: Review of Regionalization Methods. *Journal of Hydrologic Engineering*, 958-976.

Tesfaye, B. (2011). *Predicting Discharge at Ungauged Catchments Using Rainfall-Runoff Model (Case study: Omo-Gibe River Basin).* Addis Ababa University : Msc Thesis.

Tran, L. O. (2015). Linking land use/land cover with climatic and geomorphologic factors in regional mean annual streamflow models with geospatial regression approach. *Progress in Physical Geography*, 1-17.

Vaze J., J. P. (2012). Guidelines for rainfall-runoff modelling: Towards best practice model application,. *eWater Cooperative Research Centre(Vol. I).*

Ven Te Chow, D. R. (1988). *Applied Hydrology.* New York: McGrawHill Book Company.

Vogel, R. M. (1999). Regional Regression Models of Annual Streamflow for the United States. *Journal of Irrigation and Drainage Engineering*, 148-157.

Waylen .P. and M. Woo. (1982). Prediction of Mean Annual Flow in the Fraser River catchments, British Columbia. . *ResearchGate*.

Wijngaard, J. B. (2003). Homogeneity of 20th century European daily temperature and precipitation series. *nternational Journal of Climatology*, 679-692.

Yu, P.-S. H.-P.-T.-C. (2009). Estimation of Design Flow in Ungauged Basins by Regionalization. 2494-2999.

Yusuf Javeed and Apoorva KV. (2015). Flow Regionalization Under Limited Data Availability - Application of IHACRES in the Western Ghats. *ScienceDirect*, 933-941.

Zelelew, M. B. (2012). *Improving Runoff Estimatio at Ungauged Catchments. Dissertation* . Trondheim: Norwegian University of Science and Technology.

APPENDIX - A Rain gauge priority and regression used to fill the gap in rain gauge stations

Rain gauge	Missed data	Regression Eq.	Rain gauge Priority for infilling			
		Pt = α*Pd	1	2	3	4
Agaro	14.84%	Donor gauge	Choche	Gembi	Babu	
		α	0.8154	0.9388	0.8129	
		R^2	0.79	0.63	0.69	
Alibo	9.60%	Donor gauge	Jermet	Kiramu	Neshi	Kokeffe
		α	0.888	0.6283	0.6955	0.4636
		R^2	0.79	0.69	0.6	0.6
Ancharo	4.04%	Donor gauge	Tebasit	Dessie Zuria	Combolcha	
		α	0.7605	0.8745	1.1034	
		R^2	0.83	0.83	0.85	
Arjo	9.85%	Donor gauge	Dedessa river	Getema	Kone	
		α	1.2346	1.3154	0.9242	
		R^2	0.6	0.7	0.64	
Babu	10.86%	Donor gauge	Somodo	Gembe	Choche	Agaro
		α	0.9121	1.0755	0.9287	1.0957
		R^2	0.75	0.66	0.74	0.69
Combolcha	16.92%	Donor gauge	Fincha	Gebete	Homi	Neshi
		α	0.7755	1.0133	0.9381	0.8541
		R^2	0.67	0.67	0.71	0.6
Debre Berhan	4.80%	Donor gauge	Kotu	Gudoberet	Deneba	Debre Sina
		α	1.0877	0.8318	0.6472	0.5256
		R^2	0.93	0.72	0.77	0.66

Debre Markos	5.56%	Donor gauge	Aneded	Yejube	Lumame	Elias
		α	0.9865	0.8393	0.9512	0.6891
		R^2	0.77	0.71	0.78	0.76
Dejen	2.02%	Donor gauge	Abbay Sheleko	Yetemen	Lumame	
		α	1.0652	0.8423	1.0182	
		R^2	0.72	0.58	0.74	
Derba	7.32%	Donor gauge	Fital	Muke Turi	Enchini	
		α	1.1915	1.1606	1.0004	
		R^2	0.56	0.59	0.55	
Enchini	7.07%	Donor gauge	Muger	Jeldu	Derba	Shekute
		α	0.7514	0.6929	0.7206	0.6649
		R^2	0.75	0.47	0.44	0.57
Enfranze	17.42%	Donor gauge	Maksegnit	AddisZemen P.S		
		α	0.9516	0.6967		
		R^2	0.74	0.74		
Felege Berhan	2.72%	Donor gauge	DebreWork	MertoLemariam	Gundowoin	
		α	1.2651	1.0654	0.8069	
		R^2	0.66	0.6	0.77	
Filkilik	9.34%	Donor gauge	Gohatsion	Abbay Sheleko		
		α	0.9406	0.9197		
		R^2	0.78	0.65		
Gebre Guracha	4.80%	Donor gauge	Degem	Gohatsion		
		α	1.0291	1.0113		
		R^2	0.87	0.8		
Gedo	5.30%	Donor gauge	TokeErenso	Goben		
		α	0.6123	0.5913		

		R^2	0.64	0.65		
Gohatsion	2.78%	Donor gauge	Filiklik	G/Guracha		
		α	0.9186	0.8724		
		R^2	0.76	0.8		
Gonder A.P.	6.31%	Donor gauge	Chandiba	Ayemba	Maksegnit	
		α	0.8321	0.8728	0.9509	
		R^2	0.64	0.64	0.79	
Gundo Woin	8.08%	Donor gauge	Merto Lemariam	Felege Berhan		
		α	1.2229	1.0784		
		R^2	0.57	0.78		
Hareto	18.94%	Donor gauge	Goben	Fincha	Combolcha	
		α	0.926	0.636	0.6754	
		R^2	0.68	0.67	0.54	
Kidamaja	15.65%	Donor gauge	Askuna	Enjibara	Gimijabet	
		α	0.888	0.9453	1.0086	
		R^2	0.83	0.88	0.8	
Kiltukara	6.06%	Donor gauge	Mendi	WereJiru		
		α	1.1109	1.3831		
		R^2	0.64	0.59		
Kone	7.83%	Donor gauge	Mendi	WereJiru		
		α	0.9158	1.241		
		R^2	0.69	0.64		
Makisegnet	3.54%	Donor gauge	Chandiba	Enfranze	Gonder AP	
		α	0.7551	0.8628	0.923	
		R^2	0.55	0.73	0.82	
Mendi	20.20%	Donor gauge	Kiltukara	Bambase		

		α	0.7042	1.1483		
		R^2	0.61	0.55		
Nekemt	9.09%	Donor gauge	Muletadiga	Getema	Arjo	Sibusire
		α	1.1942	1.4642	1.0154	1.4564
		R^2	0.82	0.77	0.76	0.82
Neshi	6.06%	Donor gauge	Homi	Fincha	Alibo	Jermet
		α	1.0439	0.8295	1.1582	1.1788
		R^2	0.72	0.62	0.67	0.69
Sibusire	7.32%	Donor gauge	Kone	Getema	Muletadiga	
		α	0.6492	0.9545	0.7556	
		R^2	0.61	0.78	0.75	
Tikure Enchine	4.55%	Donor gauge	Gimbi Bila	Tok Ereneso		
		α	1.1716	1.1113		
		R^2	0.82	0.63		
Toke Erneso	0.75%	Donor gauge	Gedo	Tikure Enchine		
		α	1.344	0.7421		
		R^2	0.64	0.64		
Woreta	6.06%	Donor gauge	Wanzaye	Amed Ber	AddisZemen PS	
		α	0.8935	0.9652	0.8472	
		R^2	0.8	0.89	0.71	
Yetmen	7.32%	Donor gauge	Dejen	Lumamme	D/Markos	
		α	0.9078	0.9773	1.1019	
		R^2	0.62	0.69	0.66	

APPENDIX – B Tested forms of the equations for rainfall-runoff relationships

Annual flow versus Annual rainfall relationship in the form of linear and non-linear empirical equations.

Station	Forms	α	β	γ	R2	R2 Adj	Ns	Remark
Abbay Kessi	Power	2.50E-04	2.54		0.54	0.52	0.57	
	Linear	16.11	0		0.11	0.06	0.39	
	Transformed	0.0000139	2.94		0.69	0.67	0.69	Selected
	Polynomial	0.012	0	0	0.37	0.33	0.55	
Abbay@BDR	Power W cof	251.86	0.46	988.08	0.21	0.17	0.21	
	Power	0.076	1.5		0.19	0.16	0.2	
	Linear	2.85	0		0.09	0.05	0.18	
	Transformed	0.02	1.67		0.21	0.17	0.21	
	Polynomial	0.001	1.39	0	0.19	0.16	0.2	
Andassa	Power W cof	0.31	1.02	756.3	0.48	0.46	0.48	
	Power	0.00015	1.97		0.47	0.44	0.48	
	Linear	0.19	4.60E-07		0.12	0.08	0.37	
	Transformed	7.60E-05	2.06		0.5	0.47	0.5	Selected
	Polynomial	0.00012	0.0015	1.01	0.48	0.45	0.48	
Anger	Power W cof	3.61E-07	3.27	622.99	0.66	0.63	0.64	Selected
	Power	1.19E-06	2.88		0.18	0.09	0.48	
	Linear	1.211	0		0.02	-0.08	0.21	
	Transformed	1.00E-03	1.93		0.1	-0.05	0.39	
	Polynomial	0.0008	0	0.05	0.1	-0.01	0.38	

Ardy	Power W cof	34.73	0.27	0	0.04	-0.04	0.04	
	Power	34.76	0.27		0.04	-0.04	0.04	
	Linear	0.04	200.84		0.03	-0.05	0.03	
	Transformed	14.74	0.38		0.08	0	0.08	
	Polynomial	0	0.04	200.85	0.03	-0.05	0.03	
Azuari	Power W cof	71.93	0.17	1238.06	0.93	0.91	0.93	Selected
	Power	6.20E-06	2.36		0.66	0.59	0.73	
	Linear	0.119	0		0.12	-0.05	0.48	
	Transformed	3.10E-07	2.77		0.69	0.63	0.74	
	Polynomial	0	0.119	0	0.12	-0.05	0.48	
Bello	Power W cof	0.0026	1.55	85.35	0.51	0.49	0.51	
	Power	0.0012	1.64		0.51	0.49	0.51	Selected
	Linear	0.013	0		0.19	0.15	0.43	
	Transformed	0.0011	1.65		0.45	0.42	0.45	
	Polynomial	7.20E-05	0.0001	34.77	0.51	0.48	0.51	
Beress	Power	5.9	0.39		0.02	-0.01	0.02	
	Linear	0.03	57.14		0.02	-0.02	0.02	
	Transformed	2.42	0.51		3.00E-02	-0.004	0.03	
	Polynomial	0.00E+00	0.076	0.53	0.12	0.08	-0.02	
Birr	Power W cof	0.00000685	2.61	517.88	0.77	0.76	0.77	Selected
	Power	3.77E-09	3.49		0.63	0.6	0.76	
	Linear	0.346	0		0.06	-0.01	0.35	
	Transformed	0.000161	2.04		0.23	0.17	0.61	
	Polynomial	2.20E-04	0.00	0	0.22	0.15	0.59	

Chemoga	Power	2.03E+00	0.6		0.03	-0.04	0.03	
	Linear	0.06	61.3		0.03	-0.04	0.03	
	Transformed	1.05E+00	0.68		0.04	-0.03	0.04	
	Polynomial	0.00E+00	1.08E-01	1.11	0.09	0.03	0.02	
Dedissa Arjo	Power	22.26	0.67		0.38	0.3	0.37	
	Linear	1.315	1060.63		0.37	0.3	0.37	
	Transformed	30.97	0.63		0.33	0.24	0.33	
	Polynomial	0	148	797.89	0.47	0.41	0.37	
	Linear	0.1145	0		0.19	0.13	0.42	
	Transformed	0.0013	1.5994		0.47	0.43	0.48	
	Polynomial	5.36E-05	0	35.9900	0.5	0.46	0.5	Selected
Dura	Power W cof	1.24	0.86		0.54	0.49	0.54	Selected
	Power	0.036	1.28		0.53	0.47	0.54	
	Linear	0.306	0		0.33	0.26	0.51	
	Transformed	0.03	1.305		0.51	0.47	0.52	
	Polynomial	0	0.28	48.794	0.28	0.21	0.49	
Gilgel Abbay	Power W cof	762.78	0.14		0.51	0.48	0.51	Selected
	Power	9.09	0.71		0.44	0.4	0.46	
	Linear	0.734	476.83		0.45	0.41	0.45	
	Transformed	5.72	0.78		0.48	0.45	0.48	
	Polynomial	3.60E-06	1	10.04	0.86	0.85	0.38	

Guder	Power	0.356	0.957		0.21	0.16	0.2	
	Linear	0.25	16.83		0.2	0.16	0.2	
	Transformed	0.49	0.91		0.19	0.15	0.19	
	Polynomial	0.00E+00	0.25	9.99	0.21	0.17	0.2	
Gudla	Power W cof	2.00E-02	1.33	0	0.25	0.17	0.24	
	Power	2.00E-02	1.33		0.25	0.17	0.24	
	Linear	2.40E-01	0		0.14	0.05	0.23	
	Transformed	3.00E-02	1.27		0.23	0.15	0.23	
	Polynomial	1.00E-04	0.0001	120.2	0.25	0.16	0.25	
Gumara	Power	0.012	1.61		0.42	0.4	0.43	
	Linear	0.93	0		0.17	0.14	0.37	
	Transformed	0.0052	1.72		0.46	0.44	0.46	
	Polynomial	4.40E-04	0.35	0	0.42	0.4	0.43	
Hoha	Power	2.13E+01	0.13		0	-0.06	0	
	Linear	0.0045	48.76		0	-0.09	0	
	Transformed	24.57	0.1		0	-0.09	0	
	Polynomial	0	3.20E-02	1.08	0.1	0.01	-0.07	
Jedeb	Power W cof	1.90E-06	2.68	381.77	0.72	0.69	0.69	Selected
	Power	3.56E-08	3.11		0.54	0.49	0.68	
	Linear	0.17	0		0.06	-0.04	0.33	
	Transformed	9.02E-08	2.98		0.05	0.44	0.63	
	Polynomial	1.00E-06	0.17	0	0.06	-0.03	0.33	

Koga	Power	5.50E-06	2.31		0.2	0.16	0.23	
	Linear	0.09	0		0.04	-0.01	0.15	
	Transformed	1.70E-06	2.47		0.22	0.18	0.25	
	Polynomial	5.50E-05	0	0	0.15	0.11	0.22	
Little Fettam	Power W cof	23.66	4.40E-01	0	0.06	-0.02	0.06	
	Power	2.37E+01	0.44		0.06	-0.02	0.06	
	Linear	1.60E-01	334.24		0.06	-0.02	0.06	
	Transformed	3.38E+01	0.39		0.04	-0.04	0.04	
	Polynomial	0.00E+00	0.16	334.31	0.06	-0.02	0.06	
Megech	Power W cof	10.8	0.49	916.21	0.57	0.54	0.57	Selected
	Power	1.70E-05	2.27		0.49	0.46	0.54	
	Linear	0.1434	0		0.11	0.05	0.38	
	Transformed	6.60E-06	2.4		0.51	0.48	0.62	
	Polynomial	0.0001	0	0.005	0.39	0.35	0.53	
Mugher	Power	0.049	1.204		0.5	0.47	0.5	Selected
	Linear	0.2120	0.0000		0.34	0.3	0.48	
	Transformed	0.0480	1.2100		0.5	0.47	0.51	
	Polynomial	3.02E-05	0.1720	0.0053	0.48	0.45	0.49	
Neshi	Power	1.21E-07	2.9280		0.69	0.67	0.7	Selected
	Linear	0.1500	0		0.08	0.03	0.38	
	Transformed	0.0002	1.928		0.28	0.24	0.59	
	Polynomial	1.05E-04	0	0.0050	0.32	0.28	0.62	

RobiGumero	Power	2.35E-03	1.63		0.6	0.57	0.59	
	Linear	0.2042	0		0.26	0.21	0.52	
	Transformed	0.0018	1.66		0.63	0.61	0.63	Selected
	Polynomial	1.05E-04	0.073	0	0.59	0.57	0.6	
Shina	Power W cof	4.13E-08	3.120	564.85	0.73	0.71	0.58	Selected
	Power	1.03E-09	3.440		0.41	0.35	0.54	
	Linear	6.00E-02	0.000		0.04	-0.05	0.22	
	Transformed	7.67E-07	2.530		0.37	0.31	0.37	
	Polynomial	9.50E-07	0.064	0	0.05	-0.04	0.22	
Temecha	Power	0.0000558	2.1		0.62	0.58	0.62	
	Linear	0.17	0		0.15	0.04	0.45	
	Transformed	0.0000808	2.05		0.57	0.51	0.57	Selected
	Polynomial	0.000115	0	0	0.57	0.51	0.62	
Wenka	Power	7.90E-04	1.6		0.15	0.06	0.15	
	Linear	0.0580	0		0.06	-0.04	0.13	
	Transformed	2.13	0.370		0.47	0.41	0.53	
	Polynomial	3.46E-05	0	15.1800	0.15	0.06	0.15	

Appendix - C Homogeneity test

Station Name	Homogeneity tests p value				Class
	Pettitt	SNHT	Buishand	Von Neumann	
Agaro	0.91	0.67	0.42	0.43	Useful
Alibo	0.25	0.90	0.83	0.59	Useful
Ancharo	0.03	0.99	0.97	0.68	Useful
Arjo	0.00	0.00	< 0.0001	0.00	Suspect
Babu	0.40	0.25	0.27	0.11	Useful
Combolcha	0.99	0.71	0.41	0.88	Useful
DebreBerhan	0.04	0.04	0.02	0.31	Suspect
DebreMarkos	0.39	0.30	0.09	0.28	Useful
Dejen	0.49	0.26	0.11	< 0.0001	Useful
Derba	0.01	0.10	0.01	0.02	Suspect
Enchini	0.72	0.21	0.15	0.04	Useful
Enfranze	0.97	0.36	0.66	0.87	Useful
Felege Berhan	0.38	0.050	0.17	0.051	Useful
Filkilik	0.57	0.35	0.19	0.02	Useful
Gebre Guracha	0.81	0.17	0.22	0.21	Useful
Gedo	0.04	0.15	0.10	0.12	Useful
Gohatsion	0.59	0.19	0.33	0.24	Useful
Gonder A.P.	0.01	0.06	0.01	< 0.0001	Suspect
Gundo Woin	0.03	0.03	0.02	0.66	Suspect
Hareto	0.25	0.09	0.17	0.12	Useful
Kidamaja	0.59	0.67	0.38	0.22	Useful
Kiltukara	< 0.0001	< 0.0001	< 0.0001	< 0.0001	Suspect
Kone	0.29	0.04	0.05	0.17	Doubtful
Makisegnet	0.01	0.08	0.03	0.00	Suspect
Mendi	0.00	0.03	0.00	0.00	Suspect
Nekemt	0.31	0.05	0.12	0.12	Useful
Neshi	0.01	0.01	0.01	0.02	Suspect
Sibusire	0.12	0.43	0.91	0.05	Useful
Tikure Enchine	0.02	0.02	0.03	0.50	Useful
Toke Erneso	0.00	< 0.0001	< 0.0001	0.00	Suspect
Woreta	0.95	0.51	0.56	0.57	Useful
Yetmen	0.12	0.16	0.30	0.67	Useful